互联网内容行业从业者不容错过

厘清算法原理逻辑，实现精准内容推荐

增加用户留存，发掘内容变现途径与潜力



内容算法

把内容变成价值的效率系统

闫泽华 著

今日头条前高级运营总监 吴达 作序推荐

项亮 今日头条推荐系统架构师
张晶 腾讯资深产品经理
冯中华 网易新闻客户端前高级产品总监
周迪 新片场联合创始人
吴杰 会火CEO
林立 简书CEO
联袂力荐

中信出版集团

VISIT…

LANZAROTE Caliente.com

互联网内容行业从业者不容错过

厘清算法原理逻辑，实现精准内容推荐

增加用户留存，发掘内容变现途径与潜力



内容算法

把内容变成价值的效率系统

闫泽华 著

今日头条前高级运营总监 **吴达** **作序推荐**

项亮 今日头条推荐系统架构师	张晶 腾讯资深产品经理	冯中华 网易新闻客户端前高级产品总监
周迪 新片场联合创始人	吴杰 会火CEO	林立 简书CEO

联袂力荐

中信出版集团

本书由“ePUBw.COM”整理，ePUBw.COM 提供最新最全的优质电子书下载！！！

内容算法

——把内容变成价值的效率系统

闫泽华 著

中信出版集团

本书由“ePUBw.COM”整理，ePUBw.COM 提供最新最全的优质电子书下载！！！

目录

推荐序

自序

引言

Part 01 关于内容推荐

走近内容推荐

推荐的起点：断物识人

推荐算法：物以类聚，人以群分

从算法到应用

连接内容与人

常见的推荐问题

带着偏见看推荐

面对推荐系统

Part 02 关于自媒体

好内容为什么没人看

自媒体的数据分析

自媒体运营

自媒体变现

内容付费

本书由“ePUBw.COM”整理，ePUBw.COM 提供最新最全的优质电子书下载！！！

推荐序

我在今日头条的工作是服务于创作者群体的，其间被问到最多、也最令人头痛的有两个问题。其中，第一个问题是：你们能不能帮我推一推这篇文章？

这时，我通常会解释：很抱歉，做不到，这里的内容会由算法自动推荐。

接下来，对方几乎无一例外地会对“算法”这个词表现出极大的兴趣，并提出第二个问题：那你给我讲讲，内容推荐是怎么回事儿？

大哉问。我的天，这可不是几句话能说清楚的事儿。

为了给出足够可靠的答案，我曾专门请教推荐公司算法的总负责人，他笑眯眯地反问我：“先说说你的理解。”在听完我幼稚可笑的解释后，他把双手一摊：“你看，从字面意义上，我和你懂的也差不多啊……”

这当然是句玩笑话，真实的情况是：推荐算法实在太过复杂、精密，以至于熟稔此道的专业人士也难以向普通人言传，更别提做到简明扼要、面面俱到、通俗易懂了。

这正是闫泽华和他这本书的至为可贵之处。有句北京话叫“门儿清”，泽华对内容推荐就“门儿清”。他沉稳多思，出身于百度工程师，中途遵从内心做了产品经理，随时沉淀记录自己的判断和见解，后来又在顶级自媒体“凯叔讲故事”和算法分发的首倡者今日头条工作过。多种职业经历形成的多重视角使他能看得透彻，将准确的认知置于立体的坐标系之中，这才是超过一般从业者的地方。我相信，这些经历和特质不仅我自己不可比附，也是很多人无力效仿的。

今天，至少有4亿中国人通过算法获取内容，国内创作者群体半数以上流量来自内容推荐系统。但根据我的亲身体会，不仅多数读者和作者对这方面的情况知之甚少，甚至一些名头响亮的新闻学院教授、互联网评论家聊起这个话题，也十有八九是浮皮潦草，停留在大而化之的浅层次。泽华的这本书，技术出色而且界面友好，正适合对内容推荐有兴趣的各界人士拿来了解其机理。

一个事实是：高阶人群对内容推荐常持批判态度。以前，我和泽华讨论过原因，粗浅而言，一方面是由于算法目前仍明显有其不足和局限，另一方面是由于公众不知其运作逻辑。人这种动物，对不了解的东西总是心存恐惧。

我们共同达成的结论是：作为一个读者，只要你的感受有代表性，我们就该做出改变，就该做得更好。可是如何实现

呢？除了在日常工作中力求卓越，如今他写了这样一本小书，或多或少，也会帮助你洗脱成见。

作为一份迟来的答案，如果以后再次被人问到“内容推荐是怎么回事儿”的问题时，那么我会非常荣幸地推荐这本书给他。希望泽华累积了6年的经验和更久见识的认知能够广为人知。

吴达

今日头条前高级运营总监

本书由“ePUBw.COM”整理，ePUBw.COM 提供最新最全的优质电子书下载！！！

自序

作为一个技术出身的产品经理，我一路在互联网领域兜兜转转，却始终在与内容信息打交道：从信息搜索到内容创作，从内容推荐到知识付费；做过平台也做过自媒体，做过图文和音频，也做过视频，甚至触碰过VR（虚拟现实）。

创作的繁荣、载体的多元、分发的进化，内容信息消费在快速的发展中。我很庆幸自己能有机会亲历这个发展历程，与信息时代共同前行。

一方面，我们同信息的联系变得无比紧密。

作为消费者，我们可以随时随地获取信息，甚至有人断言，人类将不需要太多的记忆，只要依赖外脑就能够帮助我们实现信息的有效处理。

作为生产者，我们可以随时随地地创造信息。从博客到微博，再到公众号，我们不仅获得了平等发声的权利，也获得了以内容获取关注的机会，自媒体真正成为可以带来真金白银的行业。

另一方面，海量的信息也不可避免地带来种种问题。

作为消费者，既要坚守本心，经得住八卦娱乐信息的诱惑，让自己的注意力不在Kill Time（消磨时间）中虚度，还要有一双火眼金睛，能够从营销信息、标题党和虚假信息里识别真伪，不被虚假信息蒙蔽了双眼。

作为生产者，似乎更难选择。究竟是要追逐10万次以上的点击率、蹭热点，还是要坚守自己的本心？就算保证了内容上的坚持，又该如何处理运营和变现的问题？如何面对林林总总的渠道，是赚媒体广告费，还是做自营电商或知识付费，赚直营的钱……

凡有痛点，必有机会。

因为有信息查找的痛点，所以有了搜索引擎，让你可以一键直达，找到所需。

因为有信息发布的痛点，所以有了形形色色的创作工具和大大小小的创作平台，让你可以更酣畅地表达。

因为有信息阅览的痛点，所以有了推荐引擎，让人找信息变成了信息找人，被动消费诉求得到了极大的满足。

因为有信息筛选的痛点，所以有了知识付费，让你花钱买时间，让达人帮你筛选和过滤无效信息。

因为有信息消费的痛点，所以有了音频的繁荣，让你在无

暇用眼查看、用手翻阅的时候，亦能消费信息。

因为有信息表意的痛点，所以有了视频的繁荣，让你能以更丰富的方式消费和理解信息。

凡有机会，必有进化。

从信息搜索到内容推荐，是一种进化。

从免费内容到知识付费，是一种进化。

从图文自媒体到音频自媒体，再到视频自媒体，也是一种进化。

每一次进化，都会引发一群人的自问：跟得上吗？会掉队吗？

处于焦虑中的，不光有创作者，亦有平台。

载体迭代，比快更快。

分发迭代，比快更快。

与其焦虑，不如逐浪。

但是，在快速的迭代过程中，我相信一定有可以让我们慢下来的东西——那是内容的核。敬畏内容的价值，尊重它给受众带来的价值。

我带着这份敬畏，写下此书。希望在内容推荐百家争鸣的喧嚣里，给你带来一份关于内容推荐时代的参考。

它可能是错的，但一定是干的，一定是真的。

感谢你的阅读。

祝好！

闫泽华

本书由“ePUBw.COM”整理，ePUBw.COM 提供最新最全的优质电子书下载！！！

引言

也许，你的生活中曾有过这样的场景：在上下班拥挤的地铁中打开资讯应用查看新闻或逸事，午休时间拿出手机登录淘宝想要剁手买买买，坐在家中在大众点评上搜寻周末的美食聚餐去处……这一个个生活中再普通不过的场景，其实都早已处在推荐系统的掌控中：

- 内容推荐，将用户感兴趣的话题和内容呈现在用户的眼前，你看到的是明星逸事，我看到的是影片点评。
- 购物推荐，根据用户的风格和偏好进行商品推介，你看到的是潮流高跟鞋，我看到的是复古马丁靴。
- 美食推荐，结合用户的口味和住址推荐食肆去处，你看到的是日式寿喜锅，我看到的是重庆九宫格。

我们在同一时间打开了同一个应用，选择了同样的频道，甚至输入了同样的搜索词，但投其所好的结果让我们与自己更感兴趣的信息相遇。

这，就是推荐的力量！

科技赋予的神力让个体价值前所未有地凸显了出来：个性

化的推荐让人们不再被热点围绕、不再被排行榜左右，而是享受足够定制化的信息集合——资讯新闻、服饰妆容、饮食偏好，以及一个个让自己成为独一无二的个体的理由。

推荐是什么？

一言以蔽之，推荐是特定场景下人和信息更有效率的连接。

本书试图介绍的内容推荐系统，就是个性化推荐技术在内容消费场景里的一种应用，通过提供个性化的消费型内容来满足人们打发时间的需求。当你身边大量的信息获取途径，如今日头条、微博、网易新闻、手机百度，甚至Wi-Fi（无线网）万能钥匙都涌上这条内容推荐赛道的时候，我们或许有必要对内容的个性化推荐了解一二。

对用户而言，这些纷繁复杂却又引人入胜的内容，究竟是怎么来到自己眼前的？内容推荐系统缘何如此了解自己的偏好，又是否会如评论家警示的那般让自己深陷信息茧房中呢？

对自媒体而言，与订阅分发、编辑分发不同，推荐分发会给分发模型带来怎样的变数？在这个全平台分发的“战国时代”，自媒体又该如何在多平台间左右逢源，给自己的内容曝光量和品牌影响力带来大的提升？

接下来，就让我们翻开这本写给小白和自媒体的内容推荐

分发手册，一起走进内容推荐时代。

本书由“ePUBw.COM”整理，ePUBw.COM 提供最新最全的优质电子书下载！！！

Part 01
关于内容推荐

本书由“ePUBw.COM”整理，ePUBw.COM 提供最新最全的优质电子书下载！！！

走近内容推荐

推荐系统架构初探

想象一下，站在推荐系统之城前的你被抽离出一个数字的躯体。你找到一面镜子端详自己，却惊诧地发现自己的身体已被无数数据所组合：科技10%、篮球4%、历史1%、自然0.3%……

你仔细观察，甚至发现了很多连你自己都没有注意到的细节：虽然热爱旅游，但喜欢博物馆远胜过自然景观；虽然身为老饕，但喜欢西餐甜点多过中餐……

当然，你也会发现自己的身体仍然被一团团大大小小的迷雾笼罩，那是尚未被系统发掘的兴趣点，连你也禁不住好奇那些迷雾的背后究竟隐藏着什么。

定了定神，你走进了推荐系统之城。专门为你适配的一篇篇内容如一群萤火虫般朝你飞舞过来、将你围绕，你伸手点击了一条感兴趣的内容将它点亮。就在点击的那一瞬间，你身上的迷雾有一丝散去了，同时显示出了新的兴趣点——极限运动0.01%，那些原本就存在的密密麻麻的数字也有一些发生了变

化：有的权重上升，有的权重下降。每一次选择与反馈，你都在进化着自己的数字躯体。

现在，你生出了双翅，原地飞了起来，可以从高空俯瞰整个推荐系统。你看到了一个又一个数字拟态的人在城市的不同地方游走。每个人身边都围绕着许许多多的信息光点，又同其他人之间存在着若隐若现的连接。在那些信息光点中，被阅读的便会点亮，被忽略的则会变暗。每一个被点亮的光点都会分裂成更多光点，顺着人和人之间的连接飞舞到更多人的身边，就像被延续了生命一样。此起彼伏的光点明灭，共同照亮了整座推荐系统之城，让它仿佛有生命一般慢慢扩张。

上述过程听起来或许像科幻小说一样，但用来描绘推荐系统，却是再恰当不过了。

推荐是一个协作与进化的过程：

对内容而言，每一个用户既是消费者又是决策者，被认可的内容得以进一步扩散，不被认可的内容则被系统纠偏，以免给更多人造成负面影响。

对用户而言，每一次行为反馈都在不断地完善自己的数字躯体（用户画像），而系统的兴趣探索行为，也在进一步给这幅画像补充更多维度。



图1–1 用户画像

（图片制作：https://wordart.com/create）

初步了解推荐系统架构，有助于我们更好地认知这一过程：信息是如何匹配给用户的，用户的选择又是如何影响后续

信息分发过程的。

在推荐系统没有大规模应用于内容分发之前，人们更多地通过搜索的方式来获取信息。如果你对NBA（美国男子职业篮球联赛）感兴趣，或许需要时不时地在搜索引擎中查询“NBA”来了解最新资讯。

从工程角度来看，推荐系统的架构与搜索系统的架构具有一定的相似度：二者实现的都是信息与用户意图之间的匹配。搜索系统是将海量内容与用户表意明确的查询词相关联，推荐系统则是将海量内容与用户没有明确表达的偏好相关联。我们可以将推荐问题极度简化：当用户只有一个爱好“NBA”时，那么推荐引擎给用户的结果就可以近似等于搜索引擎在“NBA”这个搜索词下的结果。

首先，我们从简化的问题开始：搜索系统的架构是什么样的？

如图1–2所示，我们通常将搜索系统划分为离线和在线两部分。

其中，离线部分专注于内容的搜集和处理。搜索引擎的爬虫系统会从海量网站上抓取原始内容，并针对搜索系统的不同要求建立不同的索引体系。比如，我们假设新内容是有时效性消费价值的，那么，为了让新内容能够被更好地检索，就可以

建立专门的时效性索引用于存储几小时之内发布的内容。

索引是一个基于关键词的序列，每一个关键词对应一长串提及该关键词的内容（倒排索引）。比如，“教育”这个词命中内容1、2、3，“NBA”命中了另外一批内容1、2、4。

在线部分负责响应用户的搜索请求，完成内容的筛选和排序，并最终把结果返回给用户。

当用户输入搜索词后，系统会首先对搜索词进行分词、变换、扩充、纠错等处理过程，以便更好地理解用户的搜索意图。如图1–2所示，用户只输入了“NBA”一个关键词，系统发现“NBA”和“美职篮”是同义词，就会将其扩充，以便两个词都可以应用在索引的查询上。又如，你输入了“NB2K”，系统认为你的输入有误，就会将其纠正为“NBA2K”，并首先返回相应结果。

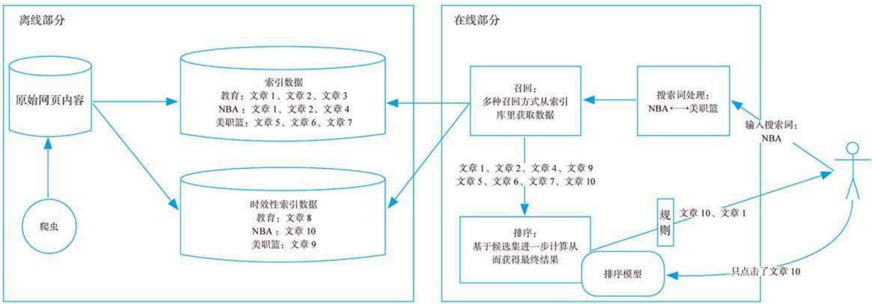


图1–2 搜索系统架构

经历完搜索词处理后，将进入召回环节。系统通过多种方

式从不同的索引数据里获得候选集合。在图1–2中，就进行了一个“2×2”的过程，两个搜索词“NBA”与“美职篮”分别查询了全量的索引数据和时效性索引数据，一共获得了8篇内容的集合。



图**1–3** 输入**“NB2K”**的显示结果

召回得到的候选集合会继续进入排序环节，通过更精细的计算模型对每一篇候选内容进行分值计算，从而获得最终结果（如图1–2中的文章10和文章1）。

在展示给用户之前，搜索结果还需要经过规则干预这一环节。规则通常服务于特定的产品目的，对计算产出的内容进行最终的调整。假如我们有这样一条“官方网站保护规则：确保所有品牌名搜索词都可以优先返回官网”，那么规则层就会在文章10和文章1之前，强行插入NBA官方网站返回给用户。

在结果展示给用户之后，用户的点击反馈会影响到排序环节的模型。在图1–4中，用户在展示给他的两篇内容中只点击了文章10，这一行为会被模型记录为统计文章10和文章1在搜

索词“NBA”下的表现情况。如果在“NBA”的搜索结果中，更多人点击了文章10，那么文章10在后续计算中的权重会升高，从而获得更靠前的展现位置（即点击调权的过程）。

以搜索系统为参考基础，我们可以更好地理解推荐系统的工作方式。

推荐系统的离线部分同样需要通过各种方式来获取待推荐的内容（如用户提交、协议同步、数据库导入等）。对内容平台来说，其内容的结构化要远胜于推荐引擎爬虫抓取的内容。

之后，离线系统依据推荐引擎对信息的不同理解维度对这些内容进行索引化处理，如话题、类目、实体词等。图1–4中展示了两个维度：分类维度和实体词维度。

在线部分，其理亦然：量化用户的请求，完成内容的筛选和排序。

推荐与搜索最大的差异在于用户表意是否明确。因此，推荐系统需要尽可能地完善用户的长期画像（对哪些类目、实体词、话题感兴趣）和短期场景（时间、地点信息），这样才能够在用户每一次请求时更好地揣摩用户当下的意图，以进行后续的内容匹配。

当用户打开内容推荐应用时，提交给系统的信息包括时间、地理位置、网络环境、手机设备型号、登录用户ID（身份

认证）等。基于用户ID，推荐系统会从数据库中取出用户的画像数据（User Profile）。如图1–4所示的推荐系统支持分类和实体词两个维度的用户画像：在分类维度，用户对体育和科技内容感兴趣；在实体词维度，用户对NBA感兴趣。基于体育、科技、NBA，推荐系统会进行下一步的内容召回和排序操作。

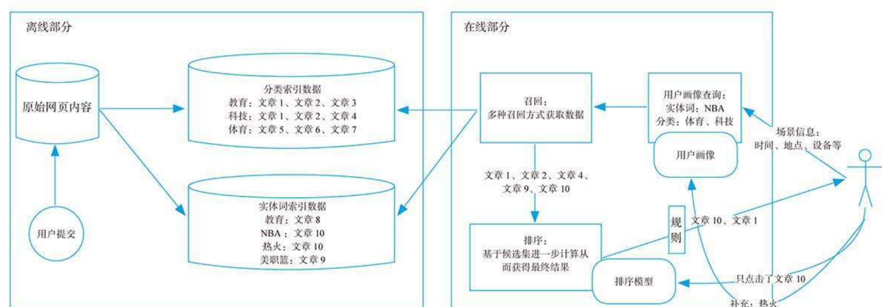


图1–4 点击调权过程图

召回和排序操作与搜索系统比较相似，系统基于类目查询和实体词查询分别获得候选内容集合。排序模块按照特定预估目标（如点击导向、互动导向、停留时长导向等）对候选集统一排序，并经过规则层的处理后最终反馈给用户。

值得注意的是，在推荐系统中，用户的行为不仅像在搜索系统中那样具有针对内容价值的群体评估意义，而且还具有针对自身画像的个体进化意义。

在群体评估意义层面，每一个读者就像是一名陪审团成员，通过自己的行为来决定某篇内容的好坏。比如某篇关于NBA的内容，如果连偏好NBA的用户都不怎么点击，那么这篇

内容在“NBA”这个实体词下的权重自然应该降低。这一部分的应用与搜索场景比较类似。

在个体进化意义层面，用户的阅读反馈行为在持续改进着自身的画像。比如，用户是热火队的粉丝，那么在行为层他就会更多地点击与热火队相关的NBA内容，系统捕捉到这一信息后，就会在他的画像中补充“热火”这个实体词，从而影响后续的推荐内容流。

YouTube和Netflix推荐架构参考

上述系统架构描述，为降低理解成本尽可能地简化了各个环节的算法复杂度。事实上，随着深度学习、神经网络的逐步工业化应用，尽管整体架构不变，但是各个环节的复杂度和算法的理解成本已经相对较高了。

以YouTube（世界上最大的视频网站）为例，2016年其已经将深度学习应用于自己的系统中，并获得了显著的效果。其系统架构如下：

YouTube的推荐架构，包含了两个神经网络，分别用于产出候选集（召回）和排序。召回层的神经网络将用户在YouTube上的历史行为作为输入内容，完成从百万量级到百量级的候选内容筛选。但其并非唯一的召回方式，YouTube历史

上应用有效的其他召回方式也同时在线上运行，如基于用户观看历史相似性、基于搜索行为、基于协同行为等方式。多种召回方式共同构建的视频候选集会被一并送进排序环节。在排序环节中，神经网络基于用户个体的特征和视频的特征对候选集进行打分，最后给用户返回得分靠前的十余个视频。

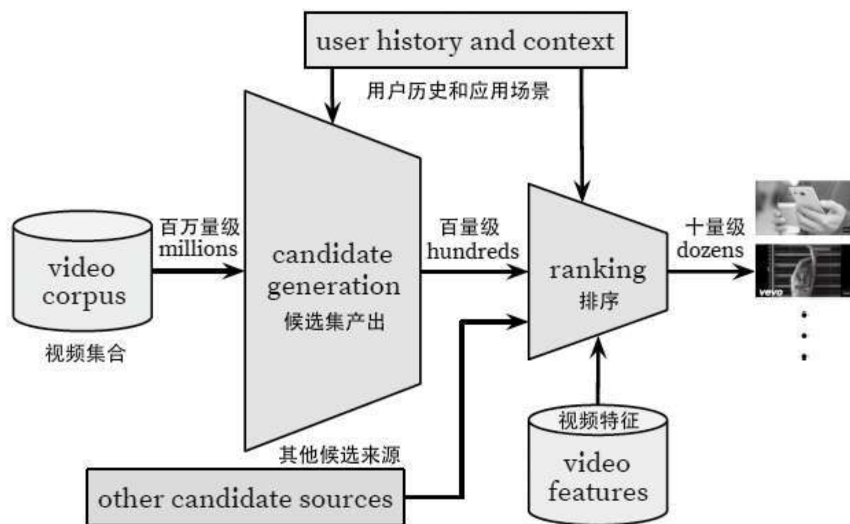


图1–5 **YouTube**内容推荐系统架构

相对YouTube较为抽象简化的数据流图，Netflix（网飞）的架构图就更复杂一些了，但整体上仍然保持了离线模型训练、在线召回排序、更新用户画像与内容画像的基础结构。系统在实时响应用户请求的在线层和负责数据处理、模型训练的离线层之间增加了NearLine（近线）层，以实现计算规模和时效性的折中。

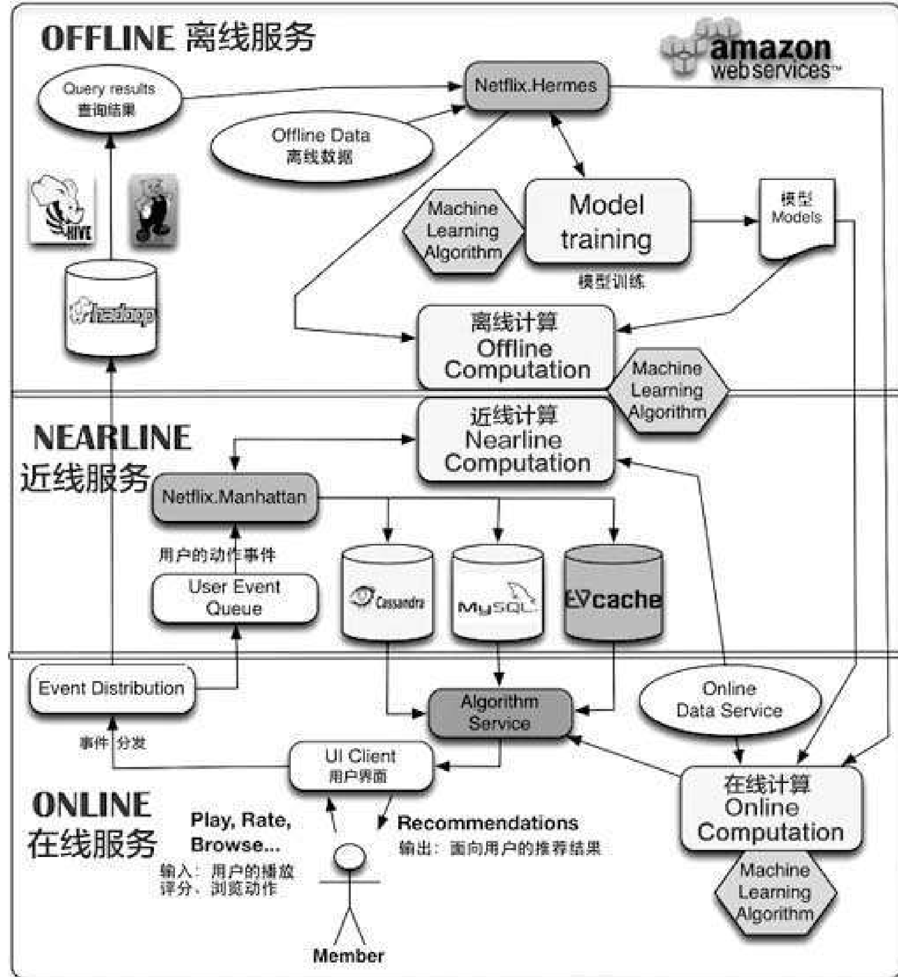


图1–6 Netflix内容推荐架构图

在数据流上，用户的所有动作（播放、评分、浏览等行为）共同构成了在线层的输入。根据业务需求，Netflix将用户行为拆分为数据和事件。其中，数据的时间敏感度更低，事件的时间敏感度更高。比如，在用户观看一部电影后，如果希望及时更新用户的推荐列表，那么这次观看动作就会被视作事件型进入NearLine层，以更快地更新用户画像数据。用户的请求经由多种召回方式从离线索引中获取结果后，经过在线的模型

排序计算，将最终结果返回给用户。

基于推荐架构的优化启示

了解了推荐系统的基础架构后，站在不同角度的我们就有了不同的优化空间和迭代导向。

用户

我们常说要把用户当作小白来看待，以不断降低用户的使用成本。但不论如何让产品普世化、小白化，每一个产品都客观存在由浅入深的功能进阶。如果作为用户的你想享有更有效率、更贴心的服务，我建议你去“训练”推荐系统。

“训练”推荐系统就请你不吝表达和互动，用你的反馈支持服务提供商和内容创作者。对于令你满意的服务和产品，登录是最好的肯定，在登录后，你的所有行为轨迹就不会丢失，在更换设备之后仍然能够获得稳定的服务体验；对于令你满意的内容，请果断地点赞、评论；对于你喜欢的作者，可以关注他的后续动态。“赠人玫瑰、手有余香”，这些典型正向反馈能够让算法更快速地收敛并确定你的喜好。对于令你厌恶的内容，也请点个“×”，明确屏蔽掉特定关键字，让系统不再做无用的探索。

在谷歌和百度已经成为互联网基础建设、自然语言搜索已经相对完善的当下，掌握搜索技巧、高效地使用搜索引擎仍然值得一书。那么，对于意图表达更隐晦、发展更不完善的内容推荐系统来说，了解推荐引擎的工作机制，更积极的参与才能够给用户自身带来更好的推荐内容。

自媒体

对内容创作者来说，只有清楚自己的内容是如何触达用户的，才能够更好地“包装”和“加工”内容。一篇内容有机会触达用户，是因为它能够被机器理解；一篇内容有机会扩散给足够多的用户，则是因为它能够收获用户的满意点击。服务于机器、服务于人，这样的内容才能够在推荐系统中获得良好的分发量。更详细的内容，会在后续章节中进行探讨。

产品经理

对产品经理来说，理解推荐系统架构有助于更好地优化产品体验、迭代分发策略，主要表现在以下几方面：

第一，完善用户画像。

通过上面的介绍我们知道，一个用户的画像越完善，就越能让一次查询的语义变得丰富，从而获得更多的候选集合，进而可能得到更好的推荐体验。

以NBA为例，每年的6月下旬到10月底，是休赛期。在这一时段内，除了偶尔哪个球星的花边消息还算得上NBA相关新闻外，基本看不到什么有关NBA的内容。对推荐系统来说，如果对用户的认知只停留在NBA这一个兴趣点上，那么这段时间显然没有办法给用户提供优质的消费内容。只有知道用户尽可能多的兴趣，内容推荐的体验才不至于因某一内容源的断供而跳崖。

完善用户画像既可以通过尽可能多的外部渠道数据塑造用户来实现，也可以借助产品设计和运营活动引导用户多沉淀行为来实现。以支付宝为例，一次过年的集五福活动，就让它收集了数以亿计的关系链数据。而紧随其后的蚂蚁森林、蚂蚁庄园等轻社交游戏，间接地丰富了用户的线下支付数据、用户的健康数据等。

第二，优化信息召回。

如果说用户画像的完善是一种输入层面的完善，那么召回的优化就是对输出候选集进行更好的调整。

我们可以用企业的招聘过程来类比，招聘的诉求就是给合适的岗位（输入的意图）找到合适的人（输出的信息），最终面试环节就代表了最后的排序层。在进入最终面试环节前，不同的候选人会有不同的经历：有些人会经历简历筛选、笔试筛选、群面筛选，也有些人会通过内推渠道直接跳过简历筛选和

笔试筛选环节，直接进入最终面试。

如果你想优化一次招聘，是优化候选人集合（召回），还是调整最终面试环节（排序）？

最终面试环节当然不应随便更改，作为一个排序环节，需要对所有候选人一视同仁，以保证公平性。就算你是老江湖，一样有可能在特定领域里水土不服；如果确实才华过人，初出茅庐的小将也定能过五关斩六将。

既然不应该随便调整最终面试，那怎样才能让更合适的人进入最终面试环节呢？

这就需要设置各种筛选维度，从不同角度进行召回：如果你是行业领军企业中的核心人员，那么基于行业口碑的背书，完全不需要笔试环节就可以直接进入面试；如果你是初出茅庐的新手，但是学习背景同岗位高度匹配，那么简历上陈述的信息就会帮助你通过简历筛选进入笔试环节。只有不同的筛选方式共同作用，才能做到查遗补缺，找到更好的候选人集合。

这就是为什么我更建议大家修改召回模块而非排序模块的原因。修改召回模块扩充候选集合，能够让我们拥有更多可能性；而主观修改排序模块则极有可能让我们损失公平，降低效率。

第三，完善规则系统，优化用户使用体验。

由于规则是在排序环节之后生效的，对系统的影响最大，因此它是产品经理最强意志的体现。这主要体现在以下两方面：

一方面，规则是最快的上线生效途径，可以用于纠偏、提权等操作。举个例子，在《中国有嘻哈》开播之前，大众是不知道红花会是什么的。彼时你搜索红花会，搜索引擎会告知你它是金庸先生《书剑恩仇录》里的一个江湖组织。而当《中国有嘻哈》开播后，所有人突然开始搜索红花会了，系统的滞后让它无法快速理解用户的真实意图。这时，产品经理就该进行规则干预了，标注红花会是一个嘻哈团体。

另一方面，我们需要认识到的是：短期的干预是应该逐步被长期的机制所替换的。规则就像打补丁，短期打补丁能够遮住窟窿，但长期补丁叠补丁这衣服就没法穿了。太多的规则系统会严重增加系统的复杂度，降低可理解性。

同样以“红花会”为例。在这个问题上，我们应该下力气解决系统的滞后性，让它能够更快速地实现搜索意图理解的进化（比如监控全网产出的新内容，分析其中的共现词汇特点）。如果有一天，有位流量明星主演了《书剑恩仇录》，那么大众对“红花会”的认知，是否又会重回武侠小说了呢？系统里配置的规则“红花会＝嘻哈团体”是否又会变得不合时宜呢？

我个人笃信的是：从长线来看，人力所能覆盖的规则语义

和逻辑复杂度是无法胜过机器的。对于初始场景或应急情况，规则系统必不可少，是用于保证服务质量底线的措施。但随着系统的逐步复杂和进化，产品经理也应该学会退位，让机器不断逼近更好的上限。

本书由“ePUBw.COM”整理，ePUBw.COM 提供最新最全的优质电子书下载！！！

推荐的起点：断物识人

尽管推荐系统中应用了各种高深的算法、架构，但其基础原理是朴素的：更好地了解待推荐的内容，更好地了解要推荐给的人，从而更高效地完成内容与人之间的对接。断物识人，是一切推荐行为的起点。

断物

想要把内容推荐给人，首先需要充分理解内容的特点到底是什么。如果连待推荐内容的特点都说不明白，那就妄论推荐效果了。更好的提取和表达待推荐内容的特点，就是断物的意义所在。

如你所熟知的那样，断物最简单的方式就是“贴标签”。

介绍人或者物品，就是贴标签的一个典型应用。先提这样一个问题，你怎么向朋友介绍一个他一无所知的人？

如果你的朋友是程序员，你也许会说：“他是今日头条的高级工程师，负责某系统的开发。”对你的球友，你也许会讲：“他大学时期是校篮球队的，拿过两次全市冠军。”对你的

酒友，介绍方式或许又变成了："我担保他是个好酒友，有量有趣，从不冷场。"

"高级工程师""篮球""有酒量""人有趣味"都是用来描述同一个人的标签，有职业维度、有兴趣爱好。为了说明不同维度的程度，你还会辅以可量化的信息，如"负责某系统的开发""拿了市冠军"等；对于不好量化的部分，则会用自身的认知和信誉作为佐证："我担保……"

但是，这个人只有上述这些特点吗？当然不是。也许他又帅又高，有八块腹肌；也许他爱好跑步，能够跑下来马拉松全程。那为什么你在介绍的时候，只选择了他的部分标签呢？因为你考虑了受众的特点，选择了对受众来说，最容易引起共鸣、也最具有代表性的标签。给球友介绍他是高级工程师、给酒友介绍他是校篮球队的成员，尽管从输出信息的准确性上来看并无错误，但是在信息的接收效果上就产生了偏差。

标签是我们对多维事物的降维理解，抽象出事物更具有表意性、更为显著的特点。在不同的应用场景下，我们会对标签全集进行有针对性地投射，有倾向性地选用不同的标签以换取信息匹配效率最大化。

在上面这个例子中，有人会关心：对推荐对象的信息整理是否可以更好地结构化呢？比如，可以将一个人的信息划分为基本信息（性别、生日）、教育背景、职业信息（公司、职

级）、兴趣爱好等。通过更系统化的分类方式，可以完整地呈现出一个人的特点。这其实涉及“标签”和“分类”的关系。

通常来说，分类是树状的，是自上而下依次划分的。在分类体系里，每个节点都有严格的父类继承关系，在兄弟节点层都具有可以被完全枚举的属性值。比如，将人类属性基于性别划分为男女，就能够覆盖全部人类属性了。由于树状的层次结构性较好，所以在内容的分类查找领域有很多应用。比如，在淘宝购物时筛选的路径如图2–1所示，就是一种分类体系的应用场景之一。

母婴用品

宝宝奶粉	英国牛栏	英国爱他	美赞臣	雅培	婴童用品	推车	驱蚊器	婴儿床	理发器	孕产必备	内衣	内裤	喂奶枕	收腹带
	澳洲爱他	可瑞康	惠氏	贝因美		奶瓶	餐椅	背带腰凳	安全座椅		妈咪包	待产包	防辐射服	储奶袋
辅食营养	米粉	肉松	磨牙棒	果泥	纸尿裤	花王	moony	大王	帮宝适	海外直邮	海淘奶粉	海淘辅食	海淘营养	直邮花王
	益生菌	清火开胃	钙铁锌	维生素		雀氏	好奇	妈咪宝贝	安儿乐		海淘洗护	海淘奶瓶	海淘餐具	海淘孕产
童装	T恤	连衣裙	泳装	套装	童鞋	凉鞋	沙滩鞋	洞洞鞋	网鞋	亲子鞋服	母女裙	父子装	亲子T恤	亲子衬衫
	衬衫	防晒服	半身裙	短裤		学步鞋	拖鞋	帆布鞋	宝宝鞋		亲子套装	母女鞋	父子鞋	家庭鞋
玩具	沙滩戏水	早教启蒙	拼插益智	遥控模型	童车	电动车	自行车	学步车	手推车	早教启蒙	早教机	点读机	健身架	布书
	运动户外	学习爱好	卡通公仔	亲子互动		三轮车	滑板车	扭扭车	儿童轮滑		串/绕珠	床/摇铃	爬行垫	木质拼图

图**2–1** 淘宝网购物分类

应用分类时必须考虑分类权威性和信息完备性问题，避免因为子节点覆盖不全或分类错误导致的认知问题。比如，在一些内容创作平台上，用户就经常会反映自己上传作品没有合适的可选择分类的情况。通常，我们建议由专家系统进行编辑分类，这样才能够最大化分类结构的可用性。

标签是网状的，更强调表达属性关系（has a）而非继承关系（is a），只有权重大小之分，不强调包含与被包含关系。

这就使得相对于分类而言，标签的灵活性更强。在权威性方面，标签是弱化的，每个用户都可参与进来，基于自己的偏好贴标签，从而借助规模效应实现对信息表意完备性的覆盖。

由于网状结构包含树状结构，故而标签可以被应用于分类的场景。在我们设计系统时，可以先基于产品场景快速覆盖主要标签，再结合标签集合的使用频次、专家建议等因素逐步将部分入口收敛到树状的分类体系中来。

标签的产出

标签从何而来？典型的产出方式有专家系统产出（PGC）和普通网友产出（UGC）两种。

专家系统产出最经典的例子莫过于音乐推荐引擎潘多拉（Pandora）的音乐基因工程（Music Genome Project）。在这项工程中，歌曲体系被抽离出450个标签，细化到如主唱性别、电吉他失真程度、背景和声类型等。每一首歌曲都会经由工作人员耗时二三十分钟，有选择性地标注一些标签，并以0~5的分值代表这一标签的表征程度。在收听过程中，潘多拉也会向用户解释为什么会推荐这首歌曲，从中我们能够一窥其对音乐进行的标签化标注。正是这项浩瀚的工程奠定了潘多拉音乐推荐的基础，并成为它的专利法宝。（参考链接：https://en.wikipedia.org/wiki/Music_Genome_Project）

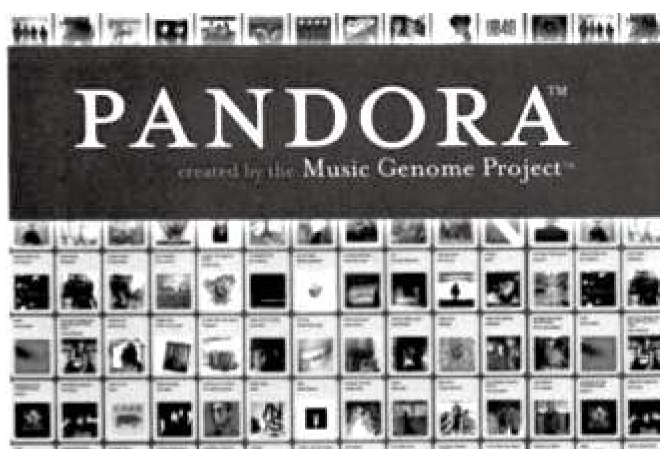


图2–2 潘多拉音乐基因工程

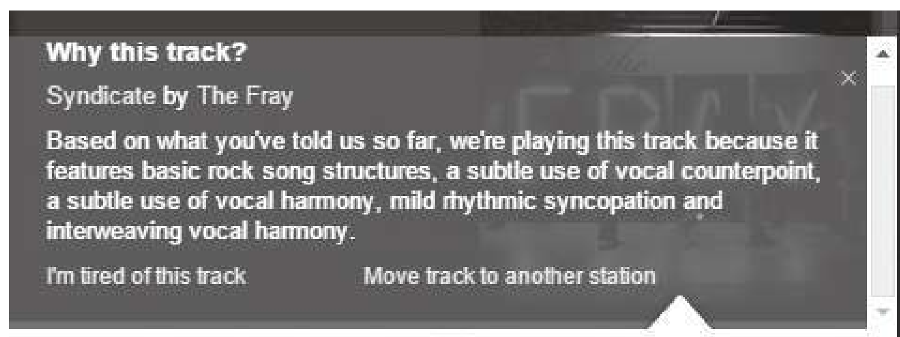


图2–3 潘多拉向用户展示的推荐理由

而在豆瓣，给音乐贴标签的活计则从专家系统产出转为了普通网友产出。群体的力量为豆瓣积累了大量具有语义表意性的标签。当然，由于用户的多样性和编辑的开放性，用户标签系统需要经过特定的清洗和归一处理。以图2–4为例，该图截自豆瓣音乐的标签系统（https://music.douban.com/tag/）。图中被框选中的部分，“摇滚”和“rock”（摇滚）表意相同为重复标签，足以说明对用户标签进行数据清洗的必要性。

风格……

OST(1264451)	pop(776382)	indie(754096)	民谣(750311)
Folk(579698)	Electronic(561917)	J-POP(484365)	流行(484298)
rock(418660)	摇滚(396218)	电影原声(357873)	JPOP(344849)
post-rock(284784)	R&B(278182)	jazz(273853)	独立音乐(217232)
中国摇滚(180125)	Brit-pop(175534)	纯音乐(173012)	britpop(160683)
古典(143883)	电子(138891)	Alternative(136900)	Metal(135367)
punk(132103)	Soundtrack(131985)	独立(123824)	classical(121137)
经典(120353)	newage(106947)	hip-hop(106868)	钢琴(106570)
原声(96164)	Post-Punk(95219)	Soul(89493)	Darkwave(88490)

图2–4 豆瓣音乐的标签系统截图

为了更好地促进用户标签的产出，我们可以借助产品层的调整和引导。豆瓣引导用户对专辑点评的界面如图2–5，它展示了层级递进的关系。

添加收藏：我听过这张唱片 x

给个评价吧?(可选): ☆☆☆☆☆ 缩起 ▲

标签(多个标签用空格分隔):

我的标签: Blues-Rock 我是歌手

常用标签: NorahJones Jazz 爵士 美国 欧美 Norah Jones 女声 Away Come

简短评论: 140

仅自己可见

分享到 我的广播 去绑定新浪微博 保存

图2–5 豆瓣音乐引导用户点评专辑界面

- 五星评价：门槛最低的评判。
- 标签输入：给出了输入框，同时也给出了候选标签集

合（包含用户自己制定的标签，以及该专辑的常见标签）。用户从已有候选集里选择的成本是显著低于输入成本的，这一设计同时实现了促进用户表达和规范化输入的效果。

- 简短评论：最高门槛的评判，深度用户沉淀内容。

当然，对断物而言，标签体系也有它力有未逮的地方，并不一定适合所有场景。

在上面的案例中，标签适合的是文字表意歧义较小、可以枚举的类型。但在某些领域，标签很难准确地表意或概括。比如，你怎样给人介绍什么是爵士乐？“余音绕梁”“节奏感十足”等，似乎都不能准确传递爵士乐的特点。又如，怎样定义一个人是帅的或漂亮的？大家各有各的标准，很难达成统一。

在这种情况下，我们需要引入聚类的方式来描述。这种方式并不是以标签词的方式来定义事物是怎样的，而是基于某一维度的特征将相关物品组成一个集合，并告诉你这个新的物品同哪个集合相似。比如，一段藏獒捕食的视频，会同动物世界里猛兽捕食的内容聚为一类，而不会同家猫捕捉老鼠的内容聚为一类。

在图2–6中，一则汽车降价促销的内容，就同多条汽车销售的内容一起被系统聚合成一类。

待聚类内容：英朗**XT**降**2.6**万 贷款购车免息送油卡

马自达CX–7现金最高直降2万元

别克君越现金最高直降3.6万元 现车销售

沃尔沃V40现金最高直降5万元 现车销售

揽胜极光促销，置换再享优惠

图**2–6** 内容聚类示例

识人

与“断物”相对应，我们可以直观地把“识人”理解为给目标用户贴标签的过程，通过标签来描述一个用户的特征集合。

图2–7就是一个用户在豆瓣FM（豆瓣音乐的频道之一）上的音乐基因名片（用户画像）。如页面中的介绍，豆瓣FM基于用户的红心歌曲、专辑收藏记录进行计算，从而得出记录用户音乐喜好的歌手和风格列表的标签云，标签的尺寸越大代表用户对这个类型的音乐越感兴趣（如果你也想测试一下自己的音乐基因，请访问：https://douban.fm/musicgene）。

豆瓣FM
doubban.fm

音乐基因名片



图**2–7** 某用户在豆瓣**FM**上的音乐基因名片

通常，用户画像有三类主要应用场景：

（1）精准广告营销。用户画像对广告营销来说是最典型的应用场景。当给用户打上各种维度的标签之后，广告主就能够借助这些标签来圈定用户，以便更有效率地触达目标人群。我们以脸谱网的广告投放系统为例，广告主可以选择地域范围、年龄范围、性别、兴趣标签等。通过选择这一系列标签，系统也会实时反馈给广告主目前框定的候选人数。

（2）行业研究。借助用户画像，我们可以了解不同行业的动态进展。比如，90后、00后的购物娱乐消费分析，不同地域用户的消费差异分析，特定行业中用户的消费特点，等等。通常，平台会定期发布此类报告，帮助内外部人士更好地了解

细分领域的最新特点。比如，在腾讯大数据（http://bigdata.qq.com/reports）上，你就可以查看到过往发布的行业研究报告。不久前，引发坊间争论的“成功人士用华为，普通人士用iPhone”，就出自于此。

（3）产品效率优化。信息匹配是最典型的场景，无论是国内的今日头条、淘宝，还是国外的YouTube、脸谱网，其平台都是基于用户的画像信息来优化推荐排序，以实现人和信息的高效匹配，从而提升效益、降低成本的。据Netflix估算，个性化推荐系统每年为它的业务节省的费用高达10亿美元。

坊间传闻，基于用户画像的推荐排序还有一些灰色的应用：比如真假货混发，罗振宇在跨年演讲“时间的朋友”中提到，某不良商家会根据用户的收货地址来预估用户的认知水平，从而决定是发送真货、高仿，还是假货；又比如价格歧视，同样一件商品添加进购物车后，不同的用户可能会收到不同的后续反馈，价格敏感型的用户更容易收到优惠券。

我们通常将用户画像数据划分为静态和动态两类。

静态用户画像数据：用户独立于产品场景之外的属性，如性别、学历、年龄、婚育状况、常住位置、教育程度等。这些信息往往相对稳定，可通过第三方联合登录、用户表单填写等方式获取。静态数据通常具有统计性意义，比如常住位置在某高档小区的用户可能付费能力更强，女性用户可能冲动性消费

更多等。

动态用户画像数据：用户在产品场景中所产生的显式或隐式行为。显式行为包括对某篇内容点赞、评论、分享，关注了某个作者等。在众多显式行为中，由于产品场景的不同，不同行为的权重也不相同（如对于电商场景，购买的权重＞购物车的权重＞查看的权重）。隐式行为包括在某页面的停留时长、用户的操作行为轨迹等。通常，显式行为的权重要高于隐式行为，但是由于显式行为更稀疏，所以需要隐式行为来补充验证。

以下是一些常见的静态或动态用户画像数据：

位置信息。　不同产品对常住位置有不同的应用方式：推荐本地新闻和特定水电暖通知之类的内容是能够想到的最直观的应用场景；美食团购类应用，是默认基于身边位置进行推荐的。但是，当你离开常住城市进入其他城市时就会发现，应用中会增加一个旅行者模块，其提供的内容更多考虑知名度而非距离因素。



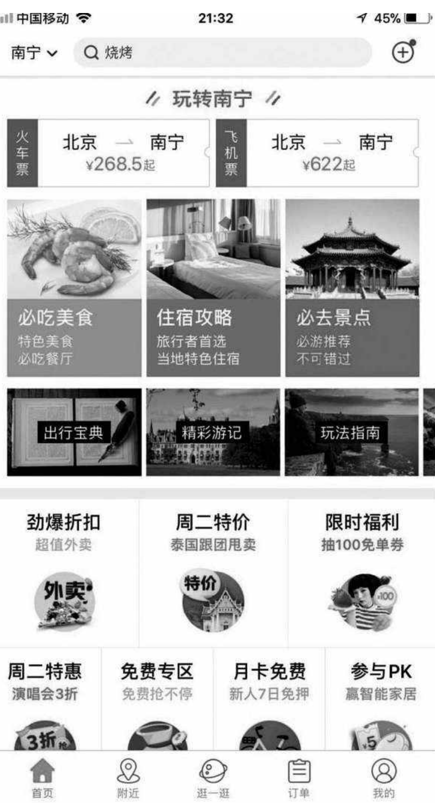


图2–8 美食团购类应用基于位置推荐示例

除了常住城市的应用外，对于朝九晚五的上班族来说，还有常住地点的概念，如高德地图和滴滴打车中提供的“家”和“公司”的选项。基于日常轨迹和常住地点，系统可以进行一些远距离的推荐，如可以在下班时段推荐家附近的饮食或尝试推荐沿途的消费地点。

搜索信息。　搜索是一个显著表明短期意图的行为，具有随时间衰减的特性。以淘宝为例，当你搜索“螃蟹”之后，那么无论是在淘宝的首页横幅广告，还是在接入了淘宝广告的应用，你都能看到关于“螃蟹”的广告。而当搜索行为过去一段时

间或是在你显式购买之后，这类广告的展现就应该降低或停止。



图2–9 淘宝基于搜索推荐示例图

评分。　评分是最常见的量化行为。如淘宝、京东等购物网站对购物行为不同维度的点评，豆瓣对书籍和电影的打分等，一般都设置为1～5分的评分机制。需要注意的是，不同用户有不同的打分偏好，比如有人习惯性好评，有人则偏严格。因此，需要将用户的历史平均分作为基准进行归一化处理，以衡量用户评分行为背后的认可度。

收藏。　收藏表达了用户对内容的偏好程度。在内容领域，具有工具性或实用性的内容通常更容易被收藏。在电商领域，收藏和加入购物车一样都是用户购买意图的体现。

分享。　除了表达喜好外，分享还传递了用户的立场和态度。比如，用户会阅读、评论很多热门的内容，但在转发的操作上是审慎的。转发到微博或微信朋友圈的动作代表了用户在用自己的社会身份扩散内容。从某种角度而言，转发的肯定意义比收藏、评论等行为的意义要大。

评论。　评论代表了参与度，但不一定明确地关联到态度的好恶。对评论的处理需要进一步进行文本分析，以获取用户的表意性和情感倾向性。在产品层上，淘宝的宝贝印象算是一个典型应用，通过抽离出用户的共同评语来辅助新的消费者进行决策。



图2–10 淘宝网站宝贝印象截图

播放比例或播放时长。　与点击行为相比，播放时长是一个相对隐性的行为。我们通常可以用播放时长来衡量用户对特定视频点击后的消费体验。2012年，YouTube就已经调整视频排序算法，将获得观看时长更多的视频排在更优先的位置。

本书由“ePUBw.COM”整理，ePUBw.COM 提供最新最全的优质电子书下载！！！

推荐算法：物以类聚，人以群分

在理解了内容和用户的基础特征之后，我们将进一步探讨推荐算法是如何匹配用户和内容的。概括而言，这一过程所做的就是“物以类聚，人以群分”。

物以类聚：基于内容属性的相似性推荐

有了完善的物品画像，我们就可以基于其固有属性来计算物品与物品之间的相似度，从而推荐与用户历史消费相似的新物品。以内容推荐为例，其用于相似度计算的常见因素有：作者层面的相似性（基于订阅或偏好关系），内容层面的相似性（如关键词、话题、类目、聚类、标签等）。

基于内容属性的推荐方式，常见于音乐（如潘多拉）、电影和书籍（如豆瓣）的推荐场景中。以图书推荐场景为例，在豆瓣图书上，有三本书分别有如下标签：

表3–1 图书标签举例

图书	标签
《推荐系统实践》	推荐系统、数据挖掘、计算机、算法、机器学习、互联网、数据分析、人工智能
《推荐系统》	推荐系统、机器学习、数据挖掘、算法、计算机、互联网、计算机科学、数据分析
《大数据时代》	大数据、数据挖掘、计算机、互联网、大数据时代、互联网趋势、社会学、数据

从标签层面不难看出，《推荐系统实践》与《推荐系统》在标签集合层面拥有更多的相同标签，具有更高的相似度。那么，当用户A选择了《推荐系统实践》之后，系统应当优先给他推荐的是《推荐系统》，而非《大数据时代》。

如果想要进一步细化，那么可以借助TF–IDF方式给不同的标签设定权重。其基本思想是：出现频率越高的标签区分度越低，反之亦然。比如“计算机”是一个高频出现的标签，那么这个标签的区分度就没有那么高，权重较低；而“推荐系统”是一个低频出现的标签，则该标签就更具有显著性和区分度，权重较高。

基于内容属性推荐的好处在于，只依赖物品本身的特征而不依赖用户的行为，让新的物品、冷僻的物品都能得到展示的机会。其存在的问题在于，推荐质量的优劣完全依赖于特征构建的完备性，但特征构建本身是一项系统的工程，存在一定成本。在上面的例子中，如果标签词粒度不够细，不能够全面描述书的内容（比如每本书的标签词只有三个），就很难计算出

置信的相似度，达不到足够好的推荐效果。

此外，基于内容属性推荐没有考虑用户对物品的态度，用户的品位和调性很难得到诠释和表达。比如，市面上关于内容分发的书籍很多，仅从标签词上很难分辨出高下。为了在推荐中更好地引入受众反馈因素，提出了基于用户行为的“协同过滤”概念。

人以群分：基于用户行为的协同过滤

举一个生活中的场景：初次为人父母，“无证上岗”的新手爸妈们内心是激动而又惶恐的。“打听”成了他们育儿的重要法宝之一。“你家宝宝用的是什么沐浴液啊？”“你们有没有上什么早教班啊？”……我在的多个亲子群里，无时无刻不在发生这样的讨论。这样的讨论也构成大家进行后续消费决策的主要因素之一。

这种基于人和人之间的相互推荐固然是弱社交关系分发的一种形态，但促成大家进行价值信息交换和购买转化的，其实是人和人之间的相似点，即为人父母、拥有相似的价值观和消费观。

把用户的消费行为作为特征，以此进行用户相似性或物品相似性的计算，进行信息匹配，是协同过滤（Collaborative

Filtering）的基础。

协同推荐可以分为三个子类：基于物品（Item-based）的协同、基于用户（User-based）的协同和基于模型（Model-based）的协同。

基于用户的协同就契合了上面的例子，其基础思路分为两步：第一步，找到那些与你在某一方面口味相似的人群（比如你们都是新手爸妈，倾向于同一种教育理念）；第二步，将这一人群喜欢的新东西推荐给你。

如图3–1中，用户A和用户C都阅读了内容A和内容C，用户B阅读了内容B。基于阅读行为来看用户相似度，用户A和用户C更为相似。那么，当用户C阅读了新内容D之后，这一内容就应当推荐给用户A。

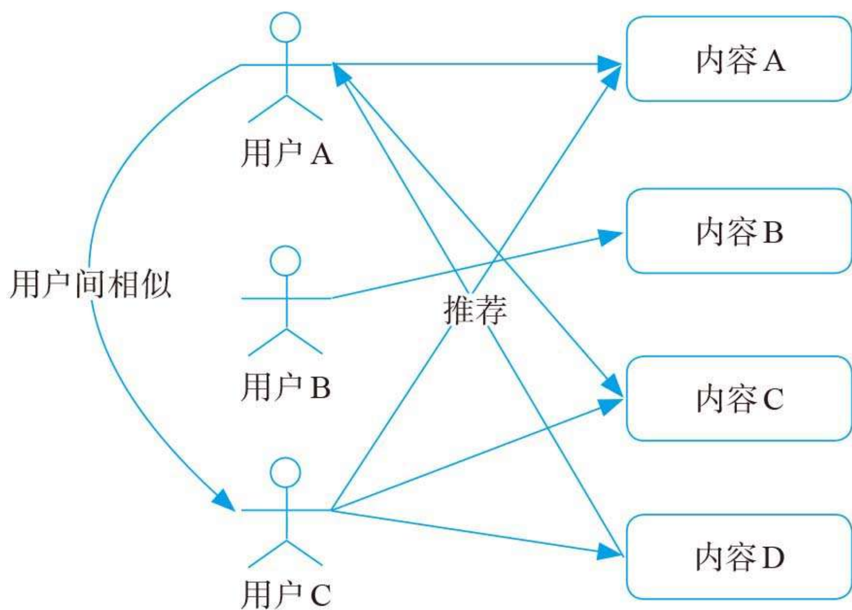


图3–1 基于用户协同示例

基于物品的协同，其推荐的基础思路是：先确定你喜欢什么物品，再找到与之相似的物品推荐给你。只是物品与物品间的相似度不是从内容属性的角度衡量的，而是从用户反馈的角度衡量的。

比如图3–2中，内容A和内容C都被用户B、C阅读，从阅读行为的角度看，两篇内容更相似。那么，当用户A阅读了内容A时，系统就选择与内容A相似的内容C，将其推荐给用户A。

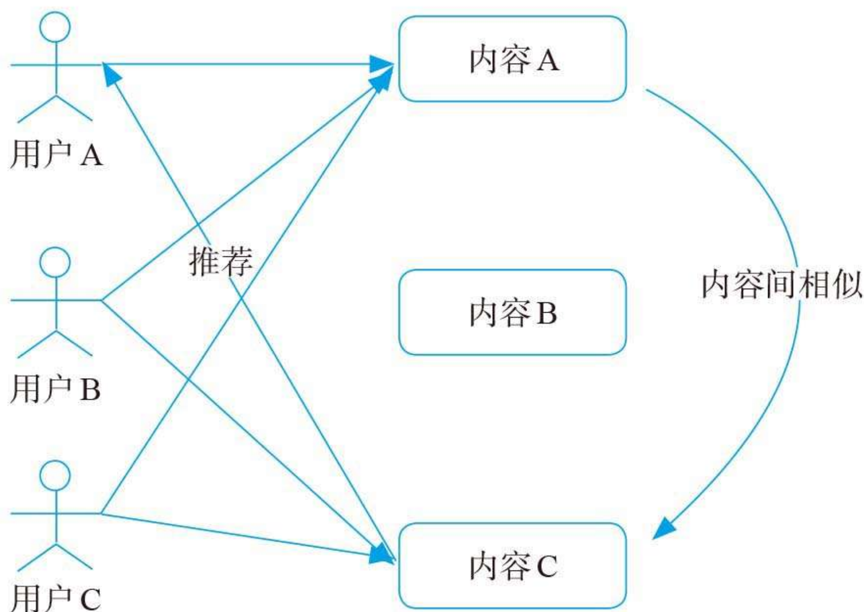


图3–2 基于物品协同示例

以书籍为例，在协同推荐的场景下，一本书的特征（标签）不再是作者、题材、领域这些静态固有属性，而是哪些用户购买了、哪些用户在购买后给了高分或低分这样的行为动作。

依赖不同的协同算法，同样的数据集合可能会得到不同的结果。表3–2中，展示了不同书籍的用户购买情况。对于用户E而言，其购买了《推荐系统实践》和《大数据时代》两本书。那下一步，我们应该给他推荐哪本书呢？

在基于物品的协同下，应该给他推荐《推荐系统》。这是因为《推荐系统》与用户E已经购买的两本书的购买用户重叠

度更高。

在基于用户的协同下，应该给他推荐《集体智慧编程》。这是因为用户E的消费历史跟用户A、B、D重叠更高、更相似，而A、B、D三位用户都购买了《集体智慧编程》。

表3–2 不同书籍的用户购买情况

书名	购买用户
《推荐系统实践》	A，B，C，D，E
《推荐系统》	A，B，C，F，G
《大数据时代》	A，B，D，E，F，G
《集体智慧编程》	A，B，D

基于用户的协同算法在1992年就已经被提出，而基于物品的协同算法直到2001年才被亚马逊提出。大家一度认为基于物品的协同要优于基于用户的协同，这是因为大型电商网站的用户数量往往远大于商品数量，且商品的更新频率相对较低，基于物品的协同能够以离线运算的方式获得更好的推荐效果。但对新闻推荐系统、社交性推荐系统等而言，其物品是海量和频繁更新的，故而基于用户的协同也有着相应的用武之地。

协同类推荐的典型应用场景，如豆瓣在书籍介绍下展示的“喜欢读×××的人也喜欢……”。

基于模型的协同，是用用户的喜好信息来训练算法模型，实时预测用户可能的点击率。比如，在Netflix的系统中就将受

限玻尔兹曼机（Restricted Boltzmann Machines，RBM）神经网络应用于协同过滤。将深度学习应用于基于模型的协同，也成了业界广泛使用的方式。











图**3–3** 协同类推荐典型应用场景示例

协同推荐是目前应用最为广泛的推荐机制，其基于用户行为的特点使我们不需要对物品或信息进行完整的标签化分析和建模，从而实现了领域无关，可以很好地发现用户的潜在兴趣偏好。考虑到协同过滤的方式依赖历史数据，新的用户和新的物品会存在冷启动的问题，将在后续的章节中进行探讨。

本书由“ePUBw.COM”整理，ePUBw.COM 提供最新最全的优质电子书下载！！！

从算法到应用

介绍了系统架构、内容画像与用户画像的建设、基础推荐算法，推荐系统的骨架算是大体搭建完了。然而，作为一套提供综合服务的推荐应用，仅有骨架是不够的，还需一副花样皮囊，通过场景划分和交互界面来实现效能最大化。

场景划分

一个复合型产品是包含多场景的，推荐算法会针对不同场景做出相应的调整和适配。参考2015年Netflix发表的论文，我们可以一窥在不同场景下Netflix的产品设计和算法应用。

垂类个性化推荐视频，如图4–1中的悬疑电影推荐、浪漫电影推荐，应用了PVR（Personalized Video Ranker）算法，针对每个用户推荐个性化的内容。这使得用户在同一类目下（如图4–1中的悬疑电影、浪漫电影）也会有不同的消费体验。

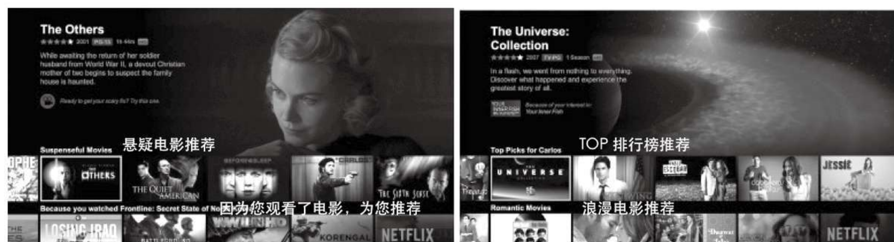


图4–1 Netflix电影推荐界面

基于用户的消费历史推荐（Because You Watched），如图4–1中“因为您观看了电影，为您推荐”，这种场景与电商网站“买了某商品的用户还购买了某某商品”很类似。推荐算法会计算两个物品之间的相似度：既可以是基于内容特征本身的相似度，也可以是基于协同过滤的相似度。通过计算相似度产出候选集，并在该集合之上进行个性化的排序和推荐。

基于Top（排名靠前）热榜的推荐，应用了Top-N Video Ranker算法。与PVR算法应用于特定垂直类目不同，Top-N Video Ranker算法选择的是多个召回序列的头部结果，优中选优，基于全品类内容提供推荐。



图4–2 基于Top热榜的推荐示例

基于时下流行的推荐，Netflix应用Trending Now算法给用

户推荐他们可能关心的短期热点，从分钟维度到天的维度不等。典型的应用场景有两种：周期性消费（节假日、重要赛事等），如情人节一定是爱情电影的消费高峰期；短期热门事件消费，如发生飓风后，有关飓风的纪录片或电影就会有一段短暂的消费高峰。对应到电商场景，这种季节性消费的特点也会比较明显，如不同季节的服饰购买需求、不同时节的农产品购买需求等。

继续观看的场景。 推荐运用Continue Watching算法，基于用户未看完的内容或正在追的剧集进行推荐。在这一场景下，并不是将用户未看完的内容以时间逆序罗列，而是综合考虑了上一次的观看时间、观看行为（暂停在了开头、结尾）等进行动态调序。此外，考虑到存在家庭账号共享的行为，算法也会进一步基于设备信息来推断当前这一用户是谁，从而展示相应的观看历史。

搜索的场景。 推荐在搜索的场景下主要用于搜索词和搜索结果的补充推荐。以图4–3左侧为例，用户搜索“usual”，但对应的影片“The Usual Suspects”系统中并未收录，从而用户得到的反馈是基于搜索意图的推荐结果；在图4–3右侧，用户输入了“fren”，除了界面右上角为搜索结果外，其余都是推荐结果，界面左下方展示了搜索词的补全，右下方展示了“French Movie”的推荐结果。

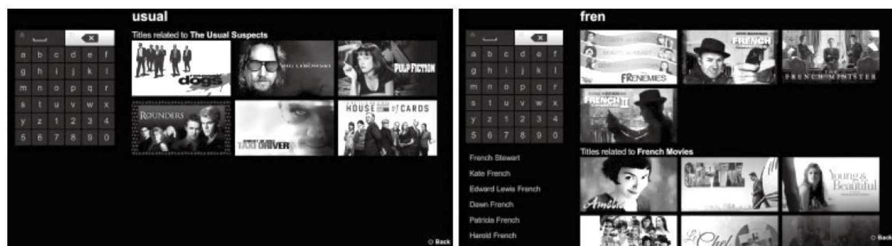


图4–3 基于搜索场景的推荐示例

交互界面

如果说推荐算法的好坏隐性地影响着用户的体验，那么交互界面的好坏正在显性地影响着用户的感知和行为。

以最常见的推荐理由为例，这一设计广泛应用于内容推荐、电商推荐的场景。如图4–4和图4–5所示，亚马逊提供的是“购买此商品的顾客也同时购买”商品推荐列表，微信看一看提供的是“好友都在读”“技术大咖在读”内容推荐标签。

从工程的角度出发，推荐理由提升了推荐系统的透明性，让用户明白为什么会推荐该种类型的内容。站在业务的角度，会更多地从促成转化入手，即什么样的推荐理由可以增加说服力，引发用户认同。

中国移动 下午1:50 27%

亚马逊 amazon.cn

与该商品相关的广告



MEGA 美高 拼搭玩具 费雪美高大袋装积木...



191 个评论

￥111.50 prime



MEGA 美高 使命召唤防空机车套装 DKX53



100 个评论

￥149.00 prime



[每天发货
孩玩具系
旅行车 ...

￥61.00

购买此商品的顾客也同时购买



LEGO 乐高 拼插类玩具 Classic经典系列 经典创意绿色底板 10700 4-99岁 积木玩具



233

￥85.00 prime



【3月新品】 LEGO 乐高 DUPLO 得宝系列 创意拼砌板 6194287 1½-5岁 积木玩具 婴幼



116

￥119.00 prime



LEGO 乐高 得宝创意拼砌系列 乐高 得宝多合一趣味桶 10572 1½-5岁 积木玩具 婴幼

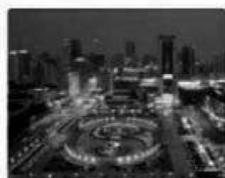
图4–4 亚马逊商品推荐列表截图

比如“技术大咖在读”这样的标签，从信息量上来看，弱化为“技术”足矣，保留此种说法的目的是为了满足用户的虚荣感。这跟服装店里导购员对你说的“美女，这件衣服好适合你啊，穿上特别显瘦”是一个道理。

另一方面，交互界面的调整也能够给系统带来更多的数据反馈和沉淀，我们通常将更被希望引导的用户动作以强形态展现（如更强化的设计、更好的入口位置），并辅以各种引导和降低门槛的措施。以大众点评为例，用户产生的商铺点评是大众点评社区内容体系中非常重要的一环，其点评界面如下：







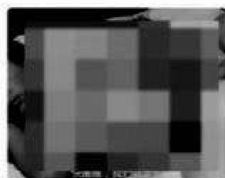




图4–5 微信看一看内容推荐截图

输入框，通过提示的方式引导用户从口味、环境、服务的角度来切入，避免无话可说的情况；图片上传，以积分和后续

的等级体系、抽奖反馈来激励用户；推荐菜，直接展示了热门菜供用户选择。各种引导和降低输入成本的措施，促进了单页面的内容积累量增加。

交互界面的迭代辅助了用户的决策，引导用户按照系统所期望的方式前进，以实现提升系统指标的最终目的。因此，对推荐系统而言，有着“交互界面 ＞ 数据 ＞ 算法”的说法。

●●●●● 中国移动 下午2:13 47%

取消 蒸汽海鲜 发表

总体 

亲，菜品的口味如何，环境怎么样，服务满意吗？

加油，还差15个字！



上传照片

3张以上有机会赢100积分

推荐菜 43道推荐菜可选>

一筐蛤蜊 皮皮虾 蒸饺 原浆生蚝 海鲜粥

海胆蒸蛋 帝王蟹 黄蚬子 蛏子王 中黄花

笔管鱼 罗氏沼虾 天鹅蛋 扇贝 捞拌四锦

人均 ¥ 请输入消费金额

@好友 

图4–6 用户产生的商铺点评界面

推荐系统评估指标

如果我们只关心推荐算法这一内核，其目标就是提高分发效率：让每一篇内容获得更多的阅读，让每一个用户更多地去点击。从算法角度看，评估指标可以大致拆分为两部分：

推荐准确度。　对应到内容推荐系统中，表现为对用户点击的预判（点击率）和对消费情况的预判（点击后的各种主动行为、停留时长）。由于准确度评估是可以复用既有数据进行离线计算的，故通常用于各种算法的迭代。比如，想设计一款新的算法以提升评论量，那么，就将历史上同一用户评论过的内容和未评论的内容输入算法，如果算法能够给前者打分更高，就意味着算法的准确率更高。

推荐覆盖度（多样性）。　好的推荐系统是能够给用户提供视野范围之外的内容的：从用户的角度看，可以评估用户的展示历史中各种题材、类目、话题的丰富程度如何，丰富度越高代表个体体验的多样性越好；从内容的角度看，可以评估有推荐展示的内容占整体内容量的比例，或整个内容分发体系的基尼系数 [1] 。

以Netflix为例，站在视频是否被播放的角度，使用了ECS（Effective Catalog Size，有效条目数量）指标来衡量系统推荐的多样性。如果系统内绝大多数的播放都来自同一部视

频，ECS指标接近1；如果系统内每部视频都有相近的播放量，ECS指标将等于影片数。更高的ECS指标代表了更好的覆盖度。

如图4–7所示，对比了应用个性化推荐算法和只采用热门排序算法的情况，两者的ECS指标相差近4倍，即系统中有更多长尾的视频都得到了有效地展示和播放。

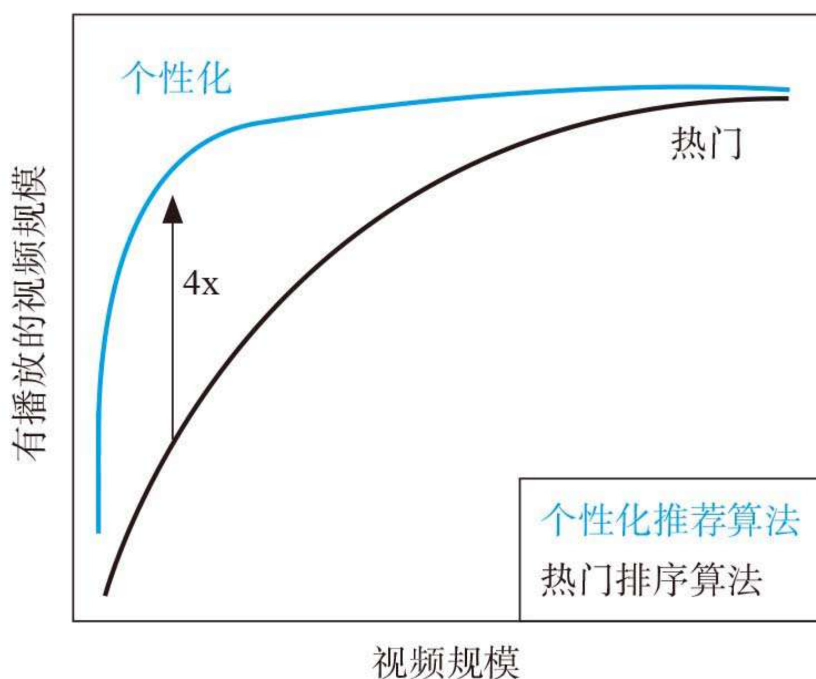


图4–7 应用个性化推荐算法与热门排序算法ECS指标对比

当然，只有一部算法引擎，推荐系统这部赛车是无法正常运转的。推荐算法应当服务于整体系统应用的目标，比如用户端的内容消费量与长期留存、作者端的活跃度等。为了这些长

线目标，会给推荐算法补充很多其他目标，这些目标可能会损失短期点击率。

同样以Netflix为例，由于其采用了付费会员制的商业模式，因此付费会员数量的多少就成了整个推荐系统的目标，并细分出新用户的付费转化、付费用户的续订、退订用户的召回等子目标。国内的视频网站，也早已从单纯追求播放量规模，转化为对视频播放量、付费用户量的多目标追求。

指标是我们对业务的抽象和预判印证，所以，客观上它是后验性的体现。如果只关注已有指标，很容易让我们忽略那些尚未被纳入指标体系之内的信息。比如：如果只有全局内容分发多样性指标而没有个体多样性指标，那么很可能出现全局内容基尼系数不高，但个体多样性体验极差的情况（如爱看足球的人看到了各种小众的足球赛事内容，但看不到自己感兴趣的内容）。又如，某些内容的点击率还不错（如蛇或其他野生动物捕食），但对敏感人群造成了极大的伤害。因此，我们需要在数据指标的基础上，增加人工评估的环节，以帮助系统更好、更快地发现潜在问题。

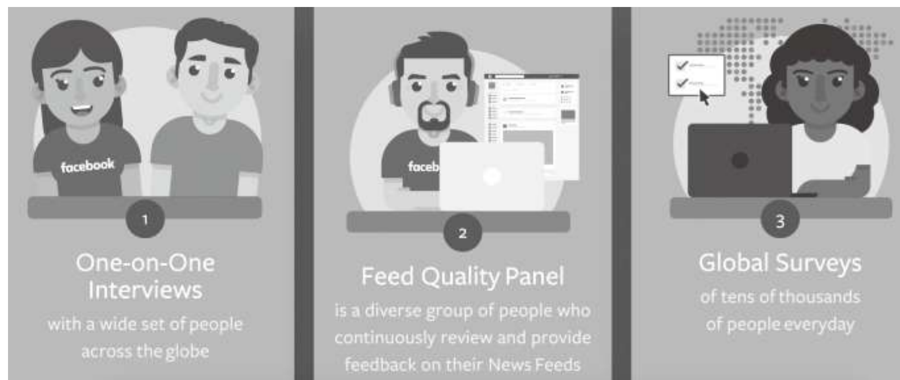


图4–8 脸谱网人工评估系统

援引公开博文。　　脸谱网建立了一套完整的人工评估系统，分为三个部分：一对一用户访谈（One-on-One Interviews），面向常住外包团队（千量级）的人工评估（Feed Quality Panel），面向普通用户（万量级）在应用内投放的问卷（Global Surveys）。其中，人工评估有多种表现形式：

（1）给出两篇内容，让用户进行点对点的对比。

（2）给单篇内容提供打分选项，建议用户从内容与自身偏好的相关性、内容的信息量等角度给予1～5分的评分。

（3）以提出开放性问题的方式来收集用户对自己信息流的反馈。

借助人工问卷反馈，我们也可以发现一些有趣的现象。比如，用户更愿意在信息流的头部看到那些自己愿意互动的内容或是那些自己觉得更重要的信息。基于此，脸谱网对排序算法

进行了调整，对用户互动预估（点赞、评论）较高的内容和用户更愿意首先看到的内容进行了提权。



图4–9 脸谱网提权内容示例

---

[1] 基尼系数：国际上用来综合考察居民内部收入分配差异状况的一个重要分析指标。——编者注

本书由“ePUBw.COM”整理，ePUBw.COM 提供最新最全的优质电子书下载！！！

连接内容与人

冷启动

作为连接内容与人的推荐系统，无时无刻不面对着增量的问题：增量的用户，增量的内容。

新的用户、新的内容对推荐系统来说都是没有过往信息量积累的、陌生的，需要通过累积一定的曝光量和互动量（阅读、分享等）来收集基础数据。这个从0到1积累基础数据的过程就是冷启动，其效果的好坏直接关系到用户端、作者端的满意度和留存率。

如果我们假设一篇新内容要经过100次阅读才能得到相对可信的内容画像，新用户同样需要完成100次阅读之后才能够建立起可用的用户画像。那么，一个最直观的问题就是：怎样达成这100次有效的阅读？

这就是冷启动面临的问题。下面我们从内容和用户两个维度来分别论述。

内容的冷启动

在推荐系统中，一篇内容借助探索性展示完成了从0到1的用户反馈积累过程。

在这个冷启动过程中，如果没能得到足够的正面用户反馈（点击行为和阅读体验），系统就会认为这篇内容是不受欢迎的，为了控制负面影响，就会逐步降低这篇内容的推荐量；反之，如果内容在冷启动过程中顺利找到了自己的目标人群，收获了很高的点击率，就有可能被推荐系统快速放大，具有了成为爆款的可能。因此，“冷启动决定一篇内容的命运”的说法丝毫不为过。

如推荐算法部分所提到的，在内容被提交进入系统伊始，由于缺乏用户行为反馈，推荐系统更依赖于内容本身的固有属性来进行冷启动。

基于内容的展示和消费，这些属性可以拆分为内容展现维度和内容消费维度。

其中，内容展现维度是列表页展现给用户的信息，如标题、封面、发布时间等。

内容消费维度包括作者层和内容层两方面：作者层内容消费维度是指，作者的粉丝群体更应该看到该作者的新内容，一个过往表现更好的作者可以得到更高的冷启动推荐量；内容层内容消费维度是指分类信息、关键词、命中的实体和话题等，

用于判断内容与用户的偏好是否匹配。

没有点击就没有阅读体验，推荐系统首先考虑的就是表现层信息与用户的匹配度。在不同的应用场景中，用户决策的依据是不一样的，信息的表现方式也是不一样的，故需要结合不同应用的展现方式进行个案分析。以内容在信息流中的展示样式为例，展示出来的信息有标题、封面、作者信息（头像、名称）、发布时间等（如图5–1）。

如果iPhone 8 真的有刘海，那App可能这样显示



武陵之荣光 220评论 8小时前



取款方式巨变！今天，微信支付宝同时宣布！

商业汇评 786评论





图**5–1** 内容在信息流中的展示示例

对新发布的内容而言，评论数和发布时间是可以忽略的，那么，展示的重点就集中在标题、封面和作者信息上：标题吸引人，封面清晰、表意明确，作者名称有权威度、与内容的领域保持一致等。

图5–1中展示的两篇内容其实都存在一些偏差：上方的内容作者名为“武陵之荣光”，对关心内容发布来源的用户来说，作者名称的权威度不够，看上去更像一个地区账号；下方的内容从作者名称看有一定可信度，但只选用了单图方式且图的表意同标题并无直接关联，不足以对用户构成有效引导。选用单图模式的另外一个问题在于，内容在信息流里的高度变矮了，不利于获取用户的注意和点击。

如前面算法部分所提到的，系统采用协同过滤的方式来对内容进行扩散，只有满意度较高的点击才算得上是对内容的有效正向反馈。冷启动阶段就是为了收获尽可能多的正向反馈，所以在预判用户有点击意图之后，还需要进一步考虑阅读体验是否同用户相匹配，即内容消费层的信息。

例如，对NBA领域的内容而言，有的是从球赛报道的角度，有的是从赛事版权的角度，还有的是从体育娱乐产业、泛娱乐经济的角度。它们在表现层展现的标题、封面的差异或许不太大，但是在内容的主题、作者或行文用字上，必然存在差异。系统通过分析作者过往的表现（如财经类作者的内容更容易倾向财经角度）、统计词频（如提及版权、分销，还是比分、MVP等关键字）等方式来预估这篇内容的分类话题属性以及用户点击后是否会满意。

就题材而言，图文载体可供系统分析的信息相对丰富。对

视频载体或短内容来说，由于文本信息匮乏，系统会更倾向于从标题、描述、作者维度来预估其点击率和内容质量。

在实践中，我们经常会碰到由于冷启动有误导致内容推荐出现偏差的情况。先来看几个例子：

- 视频标题："普京曾经实习的单位，最厉害的国家狗仔队是如何工作的"
- 短内容标题："感恩一起战斗的日子，感谢我的战友"
- 视频标题："堪比电影中出现的场景：死神来了！"

第一个视频标题可以拆分为"普京""单位、实习、工作""狗仔队"三部分。第二个短内容标题，从文本特征上来看，基本上命中"战斗""战友"。第三个视频标题，只从标题上来看，基本上命中电影《死神来了》。

但事实上，第二个短内容的场景是某个艺人在戏剧杀青后的感言。第三个视频内容描述的是车祸中某人很幸运地躲开了多车相撞，内容如图5–2所示。



图5–2 某交通事故短视频截图

上述三个内容，除了第一个内容标题命中了“普京”，使得其冷启动不至于偏差太大之外，其余两篇内容都在冷启动阶段折戟沉沙，没能获得进一步的曝光。

为了优化内容的冷启动，我们还是会从展现层和消费层分别着手：

在展现层，需要探寻面向不同载体、不同人群的，更适合的表现形式。比如：对小视频应用（抖音、快手）来说，更适合突出创作者主体，以竖屏的方式展示；对于高质量后平台强推的作品，以预览动图作为封面可以增强其吸引力。

在消费层，一方面，需要尽可能挖掘和完善不同载体的特征、权重。比如：对视频而言，长度信息、画面清晰度都可以成为补充的特征；对短内容而言，在冷启动阶段应尽量放大作者的权重，优先推荐给其粉丝和准粉丝群体，用以证明内容消费的价值。

另一方面，也需要从作者层面着手，引导作者主动完善内容信息。比如，给视频增加简介和标注，尽量选择自定义封面而非视频截图等，帮助系统获得更多可供冷启动使用的信息。业界也有公司采用专门的运营标注团队，对高质量视频进行系统化的标签完善和建设工作。

除了内容本身会影响推荐的冷启动效果之外，冷启动还客观地受到时下热点的影响。

比如，在鹿晗、关晓彤公布恋情时期，该事件构成了内容消费的流量黑洞，吸附了大多数用户的注意力。如果在这个时间点发布内容，很有可能被抢了风头，无法获得很好的冷启动效果。又如，在“双十一”期间，全民都在“剁手买买买”，整体内容消费都处于流量低谷。在这个时间段，即使内容得到了充分的冷启动依然不会有很高的最终展示量。

尽管系统在持续优化冷启动阶段的准确性，但该过程确实具有一定的偶然性。为了降低系统冷启动失败的概率，平台会提供如双标题、双封面这样脑洞大开的工具，帮助创作者获得更好的传播自己内容的机会。

用户的冷启动

在用户冷启动的场景中，推荐服务的一大目标是：用户的留存率。只有在保证用户留存的前提下，才会考量推荐的兴趣

探索效果如何，是否在有限的展示里全面探索出了用户的偏好。对于慢热型的用户，我们并不急于获悉他的方方面面，而是以留住用户为第一目的。

以知乎为例，以问答为主场景的知乎最新上线了“想法”频道，主要展示关注对象的短内容，并会穿插一些推荐性质的短内容。尽管问答主场景和“想法”频道都融合了推荐算法，但是两个场景下的冷启动目标是不一样的。

用户使用知乎是为了看答案，提问题，答问题。“想法”频道对用户而言是一个增值体验而非基础体验，做得不好，用户大不了不用这个子频道，但不至于从应用流失。所以，它可以更单纯地追求用户兴趣探索的效果。但对问答主场景而言，就需要平衡用户兴趣探索的效果（尽可能地探索出用户的兴趣点）和用户留存的情况（以高热内容留住用户）。

对产品主场景而言，用户留存一定是第一目标，只有把用户留下来才有持续探索的可能性。所以，系统在新用户的冷启动推荐上也会更偏向于重大兴趣的收敛而非发散，尽可能地追求留存率的提升。

对于冷启动的内容，我们可以基于文本分析抽离出内容的关键字、话题来建立内容画像；对于冷启动的用户，我们同样需要尽快完善信息的搜集和预处理工作，建立对用户的初步认知。对移动端产品来说，可以获取信息的途径有多种：

在设备层上，如厂商、型号都是必备信息。对安卓设备来说，某些品牌的用户在统计学上具有一定的可区分度，如锤子手机、一加手机的用户同VIVO、OPPO手机的用户就存在一定差异。

在权限层上，移动设备可补充信息的丰富程度远高于PC（个人电脑）端：设备标志信息可以帮助判断用户是否是卸载重装的老用户；地理位置权限能够帮助确定用户的当前位置、常住位置、行为轨迹。

Wi–Fi信息可以同时用于确定位置信息，并进一步将处于同一Wi–Fi环境下的用户圈定出来（通常是家庭、工作关系，如果同一Wi–Fi下的用户数量不多，那么这些用户彼此间可能有更亲密的关系）。

基于社交关系的应用还会申请读取用户的通信录信息，以此补全用户的社交关系网络，并通过用户A的手机通信录信息，反推其余手机用户的真实姓名。比如，系统已经知道了我的手机号，那么在我的朋友也上传了他们的通信录的情况下，多个朋友对我的备注信息是B，就可以大概率确认我的真实姓名是B。

由于部分安卓手机用户的权限意识较弱，很多应用还会申请如读取目录或已经安装的软件等权限，通过文件目录反推用户已经安装的应用情况，以此来确定用户的偏好信息。比如：

如果用户安装了大姨妈、蘑菇街之类的应用，则该用户极有可能是女性用户；如果用户安装了宝宝树、亲宝宝之类的应用，则有可能是有孩子的用户。

在应用层上，结合产品设计和分发渠道特点，可以获得更丰富的信息。

装机渠道：　在安卓设备上，通过不同的市场渠道、不同的广告素材转化而来的用户具有不同的特点。比如，通过金融相关素材转化的用户，显然会对金融类内容更感兴趣。在iOS（苹果公司的移动操作系统）设备上，也能通过一些类“邀请码”的设计方式来回溯用户是通过什么渠道安装的。

登录方式：　通过第三方登录能够更快速地获取用户的基本信息。国内有微博、微信的联合登录，国外有脸谱网的联合登录。其中，脸谱网支持获取的权限信息如图5–3所示，有助于分析用户的好友信息、工作信息、教育背景等。

- public_profile
- user_friends
- email
- user_about_me
- user_actions.books
- user_actions.fitness
- user_actions.music
- user_actions.news
- user_actions.video
- user_actions:{app_namespace}
- user_birthday
- user_education_history
- user_events
- user_games_activity
- user_hometown
- user_likes
- user_location
- user_managed_groups
- user_photos
- user_posts
- user_relationships
- user_relationship_details
- user_religion_politics
- user_tagged_places
- user_videos
- user_website
- user_work_history
- read_custom_friendlists
- read_insights
- read_audience_network_insights
- read_page_mailboxes
- manage_pages
- publish_pages
- publish_actions
- rsvp_event
- pages_show_list
- pages_manage_cta
- pages_manage_instant_articles
- ads_read
- ads_management
- business_management
- pages_messaging
- pages_messaging_subscriptions
- pages_messaging_payments
- pages_messaging_phone_number

图**5–3** 脸谱网支持获取的权限信息

业务设计：　　偏社交性的应用则会引导用户上传真实头像、补充身份信息等，并结合用户提交的性别、年龄等信息，基于人群上的统计意义来推荐。偏阅读性的应用会让用户手动选择自己感兴趣的领域，或通过让用户对一些有区别的内容进行打分的方式来判断用户的喜好。

最典型的应用就是主打用户主动订阅、新内容及时推送模式的“即刻”了。在启动之后，该应用会展示主题订阅页，引导用户订阅自己感兴趣的主题（如图5–4）。但这样的方式也存在一些问题：用户真的能够准确表达出自己的喜好吗？还是往往过高地估计和描述了自己的喜好情况（比如，我个人觉得自己是喜欢专业分析内容的，但显然高估了自己）？此外，这种

在主消费场景之前增加一道门槛的方式，本身也会对小白用户造成一定的影响。

App Store ●●●●● 上午10:12 51%

已有账号>

试试看关注一些感兴趣的主题

我们给你最细致、精准的信息服务。
不必担心，你随时都可以关闭推送通知。


一觉醒来世界发生了什么



今天有什么好笑的



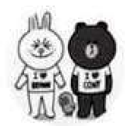

有新的热门表情包



最好看的娱乐八卦都在这里



最近朋友圈什么比较火



情侣头像精选



折扣优惠精选



《醉玲珑》更新提醒



有豆瓣9.0分以上的新电视剧



值得收的时尚单品




「阅后即瞎」更新提




热门网络梗普及


滚动查看更多!

关注几个主题试试（1/2）

图5–4 “即刻”主题订阅页

应用家族： 对BAT（百度、阿里巴巴、腾讯三大互联网公司首字母的缩写）这种有多款热门应用的公司而言，应用与应用之间的交叉验证，就已经能将未登录的设备对应上已注册的用户，从而复用已有的用户信息了。

从某种角度说，腾讯可能比你自己更了解你。腾讯的产品从最基础的社交工具QQ、微信，到新闻阅读“腾讯新闻”，再到娱乐化消费的“腾讯视频”“QQ音乐”“QQ阅读”“王者荣耀”等，不胜枚举。对腾讯系的产品而言，如果善用用户画像，怕是没什么用户算得上是“新”用户了。

面向冷启动用户的推荐，最典型的方式就是使用非个性化的热门榜单。这种方式主要满足了“新”和“热”两个特点，即使用户对推荐内容不感兴趣也不至于太过反感，待收集到一定的数据后再转换为个性化推荐。Netflix的研究也表明：在冷启动阶段新用户更倾向于热门内容，而老用户则更需要长尾内容的推荐。

进一步地，我们需要在有限的展示次数中，尽快确定用户的主要偏好。对于兴趣探测内容的选取就存在基本策略：按照覆盖面由大到小的次序，选择有区分度的内容进行用户偏好的探测，基于已有用户群体的偏好统计数据，动态调整后续用于兴趣探测的类目和内容。

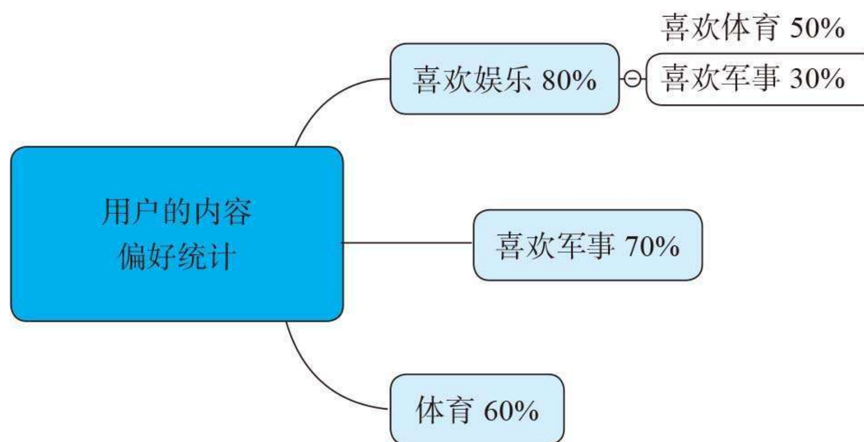


图5–5 用户内容偏好统计示例

以图5–5为例，娱乐、军事、体育是内容消费大类，可以优先展示这些类目的内容给用户。如果用户表现出了对娱乐类目的偏好，系统一方面可以纵向深挖用户对娱乐类目下各个子类的兴趣分布，另一方面也可以通过已有的“喜欢娱乐的用户是更喜欢军事，还是更喜欢体育”的统计数据，选择下一步应该优先探测军事类目还是体育类目，并沿着树状结构优先探寻用户更可能感兴趣的类目内容。

兴趣探索

承接用户冷启动的环节继续讨论。在冷启动环节里，应用推荐技术本质上还是以“留住用户”为第一要务：尽快发现用户的主要兴趣点，并推荐和这一兴趣相关的内容以换取更高的点击率。而当新用户已经成了老用户，在已经实现短期留住用户

的前提下，推荐系统需要在一定程度上牺牲短期点击率来探索用户更广泛的兴趣，从而获得用户长期留存率的提升。

兴趣探索对推荐系统而言是绕不开的一环，可以从内容供给和用户消费两个角度分别切入。

内容供给的角度

如果我们将内容推荐系统比作一家超市，那么在商品供给层面，超市需要做的就是不断地覆盖市面上值得引入的新商品，无论是自制熟食、农场直供，还是进口商品等。品类、品牌和价格区间的全覆盖能够帮助超市更好地服务于消费者。内容推荐系统在内容资源的建设上，也同样致力于提升品类、品牌的覆盖度。超市引入了新的货品或品牌供应商，会在显著位置设置堆头或展示位来帮助消费者建立认知；内容推荐系统引入了新的内容品类或创作者，同样需要让这一类内容尽快通过探索找到基础用户。

内容探索能够让我们对系统内的内容分布和内容价值体系有更好的认知。

在系统体量较大的情况下，一个确实有用的方法是“大力出奇迹”，即不考虑所有先验信息，直接将内容交给推荐引擎来判断。如果新类目的展示量过低，就调整新品类内容的加权系数，以保证内容能够展示在一定规模用户群体的信息流中，

即以规模换效果。

“姜太公钓鱼，愿者上钩”，在保证展示量的前提下，这种生推的方式通过累积用户的点击行为能够逐步挖掘出新品类的受众群体。对应到超市的场景下，或许就是你走进超市入口那个打着新品促销招牌的堆头，通过流量折损的方式让消费者知道这一新品牌，并交予消费者尝试和判断。如果你喜欢上了某种新品，后续就能够在超市对应的位置找到此类商品。

以嘻哈音乐为例，最近大火的《中国有嘻哈》节目让这一长期小众的文化迅速升温，将嘻哈音乐的内容强曝光在大众面前。不同人群的行为在大曝光量下都得以挖掘：已有认知的用户继续保持自己的偏好（喜欢或不喜欢），之前没有足够认知的用户在经过多次曝光之后，也能够慢慢建立认知，其偏好行为和点击行为趋于稳定和收敛。

“大力出奇迹，奇迹有概率。”生推的方式通常适合大的品类或内容品牌。比如，微博引入了某国际球星，那么直接送一次全局的开屏曝光也未尝不可。大品类内容是具备大规模内容生产能力和内容消费需求的，即便是生推，点击率的损失也不会太高，很可能在百万规模展示后，就能够很好地探索出这一兴趣点下的用户群体了。

但对一些小众的品类或品牌、规模较小的平台而言，采用生推方式的收敛速度就太慢了。比如，一个小众类目预估只有

千分之一的受众，那么对日活跃用户达千万的应用而言，全平台受众也只有万量级。将这个品类的内容展示一百万次，预估点击人数也只有一千，再考虑到目标受众有可能错过或误点击，其效率是很低的。

同样不适合生推的类目还有容易引起争议的、令人反感的内容品类。比如，恐怖类内容毋庸置疑是有大量拥趸的，但如果系统引入了此类内容后进行广泛的用户探索，将不出意料地收到海量的用户负面反馈。兴趣探索的前提是在尽量不引起用户反感的情况下，损失点击率进行探索，点击“不感兴趣”显然比略过不点击对用户的伤害更大。

由于生推的方式太过于简单粗暴，为了让这个过程显得稍微有策略一些，我们引入了专家系统的先验知识。基于专家的判断制定简单的人群定向规则，用规则引导新品类、新品牌的加权分发，将全人群的强展现转为特定人群的强展现，从而换取可能更高的点击率。对应超市的场景，可能是引入了速冻小龙虾，那就将其摆放在夏季热销的烤串、啤酒附近，预判消费者有共同购买的场景从而促进新品销售。

同样以嘻哈音乐为例，如果我们对嘻哈音乐的预估是年轻人、喜欢标榜个性、喜欢潮流服饰、街舞等分类内容，那么，嘻哈音乐的分发规则就会被指定为15～30岁，对潮流服饰、街舞等内容有偏好的人群，只给具有此类画像的用户进行探索分

发。

用户消费的角度

如果我们把对新类目的探索概括为“以规模换效果”，那么，对用户新兴趣的探索便可以概括为“以时间换效果”：只要用户在应用内停留的时间足够长，就有机会通过各种中低频的方式来完善用户画像。

一方面，随着用户停留时长的增加，会产生更多的主动动作，如评论、关注、搜索、调整频道顺序、主动访问不同功能页等，这些主动动作都会给用户画像补充新的信息。比如，当用户搜索特定关键词“嘻哈”时，就代表他对这类内容产生了短期兴趣。

另一方面，系统也会基于用户已有的偏好进行更广泛的兴趣探索。如在用户冷启动部分提到的，系统会基于统计学的概率——喜欢A的用户有多大概率喜欢B，喜欢A和B的用户有多大概率喜欢C——来给用户推荐新领域的内容，逐步探索和完善用户画像。理论上，如果我们划定一定比例的展示量用于新兴趣探索，那么，只要用户的停留时间够长，系统内的兴趣分类就一定能够被探索完毕。

对于用户的兴趣探索，一个值得注意的问题是：小众兴趣的探索和丢失。这个问题本身是因小众兴趣的供给不足引起

的。比如，用户的兴趣偏好是“马龙＋乒乓球”，但系统内这类内容只有100条，是个很小的兴趣点。

从探索的角度来看，内容太少会导致试错的成本太高，一次负反馈就会影响后续的探索过程：一旦因为上下文、场景等关系用户没有点击，系统就会转向其他兴趣点，从而错过此类内容的发现。

从消费的角度来看，即使系统发现了用户的小众兴趣，但是若干次刷新之后这类内容就被消费完毕了。之后，由于缺乏足够的优质内容供给，用户会在相当长的时间内没有办法触达此类内容。系统基于时间进行衰减处理，小众兴趣就会慢慢地被淡化和丢失。

应对小众兴趣，一方面需要扩充系统的资源池，让小众兴趣也有足够的内容覆盖，另一方面也需要通过产品设计鼓励用户更主动地进行强表达行为（如收藏、关注），一次关注行为显然比一次点击行为更经得起长时间的衰减。

好的推荐系统，是不会止步于已知边界的，而是会用一次次的探索去触达未知，给用户带来惊喜。

自媒体与平台

在前面，我们主要聊了内容和用户的推荐匹配过程。但这个匹配过程的前置条件是：内容从哪里来？

作者！

从推荐的角度出发，在内容冷启动的阶段，作者是具有非常高权重的因素，一个合理可信的作者价值体系，对于推荐效果的改进是大有裨益的。从内容稳定供应的角度出发，平台需要维系与作者之间积极稳定的关系，以获取持续稳定的内容；从内容生态和社会责任的角度，平台亦会积极扶持新作者、保障原创自媒体的权益。连接内容与人，我们面对的不仅是用户，还有作者。

站在平台去看自媒体，我们会分别从评估、服务、引导三个角度进行阐述。

自媒体评估

机器推荐时的作者权重计算、运营资源投入时的分配选择、新功能内测时的定向邀请，都要求我们对自媒体有相对合理可信的评估体系。虽然各大内容分发平台在内部都有着各自的细则，但其整体评估基调还是可以透过它们公布的指数信息或评级信息得以了解一二。

首先摘录头条号指数的介绍。

- 原创度：平台鼓励作者进行原创，并手动、优先在头条号平台发表内容。
- 健康度：将用户对自媒体发布文章的阅读行为进行分析和统计，用户每一次有价值的点击、停留、点赞、评论、收藏等都会为其账号加分。
- 垂直度：一方面考虑自媒体发文的专注程度，发文领域越专注其分值就越高；另一方面考虑自媒体的更新频率，发布频率高也会有助于获得更高的分数。
- 关注度：主要取决于自媒体的粉丝绝对数量、活跃情况（粉丝阅读评论情况），是对自媒体粉丝运营的考量。
- 传播度：指已发布内容的影响力，主要取决于自媒体的累计阅读量、累计播放量。

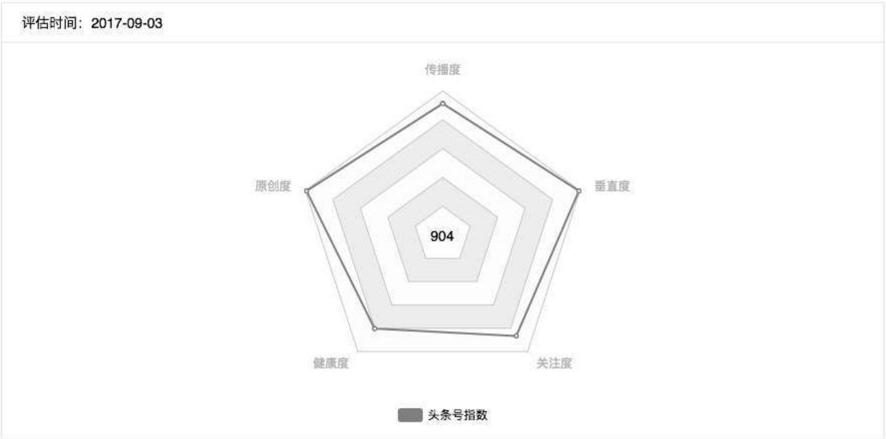


图5–6 头条号指数

通过分析其他平台的评估方式，我们不难看出，其与头条号指数的评估方式大同小异：

表5–1 其他平台评估方式

UC大鱼号	内容原创度、内容质量度、用户关注度、创作活跃度、内容垂直度
企鹅号	用户喜爱度、内容原创度、账号活跃度、内容垂直度、账号健康度
百号家	内容质量、领域专注、活跃表现、原创能力、用户喜爱

借助这些已公开的维度，我们可以抽离出平台对自媒体的评估角度：适合平台的内容价值，基于平台的持续贡献，依托于平台的作者品牌。

适合平台的内容价值： 不同的平台多多少少都会有自己的调性，外来的和尚未必合适，只有适应自家平台水土的才能够念出真经。比如，快手的“大神”到了美拍未必会卖座，美拍平台应该基于自身平台的特点来衡量作者价值，健康度衡量的就是这方面的价值。它基于内容在平台上的点击阅读表现，给出自媒体在平台上的合理价值预估，而不追求全网 PR（Publish Relations，宣传效果）意义上的“优质自媒体”。

基于平台的持续贡献： 平台希望自媒体能够持续稳定地创作专业内容，擅长体育的就持续深耕体育类目，擅长历史的就不断产出历史评说。原创度、垂直度、传播度就是对这方面进行衡量。原创度高的账号由于具备创作能力，其内容质量的稳定性、内容题材的独特性通常会高于搬运作者；垂直度高的作者更新频率稳定，产出内容领域垂直；所有基于平台的投入最后都会反馈在累计阅读量的持续增长，即传播度的维度上。

依托于平台的作者品牌： 一篇泛泛而谈的内容，你可能会阅读完毕，但绝对不会想要关注，能够让用户产出一探“结出果子的是哪棵苹果树”念头的，一定是更有价值的作者。从平台的角度看，一个具有粉丝凝聚力的自媒体账号，是有助于提升平台留存率的，因而更值得肯定。这也就是为什么在已经有了健康度这一指标可以反映单篇内容质量之后，还要引入关注度指标，用于衡量用户对作者价值的认可程度。

从实践来看，原创自媒体和垂直编辑类自媒体都具有较高的订阅价值。它们或是能够自己产出内容垂直、质量稳定的内容，或是能够持续编辑、整理同领域的内容（类似歌单），给用户带来稳定的内容消费体验。

除了已对外公布的普适标准外，伴随着发展进程，平台在不同阶段会有不同的发力重点和侧重点。

在平台初创阶段，为了追求优质内容的覆盖率，新平台会更多地对标成熟平台上表现良好的自媒体。如图5–7中，如果你拥有了开通原创权限的微信公众号或头条号，就能够直接在定位一致的大鱼号上申请原创保护。

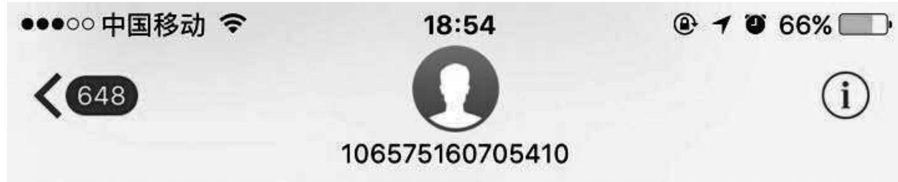


图5–7 大鱼号申请原创保护相关截图

2017年，各平台都在短视频领域发力。由于符合平台战略，在这一阶段进行短视频创作的自媒体往往更容易获得高评估价值。比如，2017年今日头条拿出了10亿元来扶持视频自媒体，大鱼号设立了专项短视频创作奖金等。对于淘宝这类以内容消费带动商品消费的平台，具有更强内容导购能力的自媒体就能够得到更多的平台青睐。

平台与自媒体之间的合作，是一个双向选择、共同进步的过程。只有更适合平台的自媒体才能够给平台带来更多的内容

消费量，也会因此收获平台更大力度的支持，从而构成一个良性循环。

自媒体服务

如果把自媒体生态比作一片森林，那么平台方更像是一个护林员，其主要职责是发现新苗、拔除毒草、辅助新苗成材。

在发现新作者方面，客观地说，相对于订阅分发和编辑分发机制而言，推荐分发机制是更加友好的。因为在这种机制下，订阅关系只是多个分发因素，一旦内容能够获得很高的点击率和良好的阅读体验，就很容易被系统加热产出爆款。只要内容质量好，不依赖平台运营的扶持也能够有持续稳定的高曝光量。

平台运营者会持续地监控平台上数据表现突出的作者，并通过人工二次校验来确认作者的内容质量，表现优秀的作者得以进入新秀榜单、获得进一步的扶持或参与平台的各类活动。目前，几家主流的平台方都逐步以动态的榜单入选、当期激励的方式替换静态签约、长期保障的方式，让优秀的新作者能够更快地涌现出来。

以2017年8月大鱼奖金的得奖情况为例，援引官方公告“奖金得主中有27%为入驻不满3个月的优质作者，而全新获奖账号（即过往8期未曾获奖）比例达到39%”。如此高的新作者获

奖比例，不难看出官方刻意引导的意图。

发现新苗子后，自然需要加大对其创作环境、创作工具的投入以促进其更好地成长，让小苗长成大树。

在版权意识逐步强化的今天，对原创生态的保护已经成了各平台的共识和基础设施。在保护平台内部创作者权益方面，微信的原创保护分享样式做得非常强。如图5–8所示，如果转载者没有获得原创者的授权，那么其发布后的阅读样式会变成如下样式，只展示被转载内容的摘要信息，用户想要阅读就必须跳转到原创者的页面，为原创者带去更多的流量。而转载者用来扩展变现的“阅读原文”链接则被替换成了官方的原创声明链接。



图5–8 微信原创保护分享样式

在降低作者维权成本方面，头条号推出的“原创保护”是一个典型案例。通过监控23家平台的内容侵权情况，系统会在6小时内完成检测，抓取出疑似侵权的链接，在作者确认后，可以选择交由维权骑士、快版权等第三方维权机构帮作者实现维权。

保护了基础原创环境之后，还需要给原创作者带来更多的

曝光量。在传统平台上，我们往往看到的是如横幅广告这样简单粗暴的强位置曝光，但在推荐主导的平台上，运营者更希望从“授人以鱼”逐步过渡到“授人以渔”，从单纯的流量分配扶持转化为更好的运营工具、更多的使用特权、更充分的沟通与指导等。

以头条号指数为例，一方面指数本身对内容推荐量有加成的作用（如头条号指数图中的实际提升推荐量），另一方面其同时关联了各种运营工具的开通权限。如：双标题双封面、热词分析、外图封面可以帮助作者获得更多的流量和更好的消费体验；商品、粉丝必见、扩展链接则可以帮助作者获得更高的扩展性和商业变现可能（如图5–9所示）。

功能状态

功能	状态	申请条件	功能说明
商品	已开通	粉丝数2000以上；头条号指数650以上	功能介绍
双标题/双封面	已开通	粉丝数2000以上；头条号指数600以上；已开通原创权限	功能介绍
粉丝必见	已开通	粉丝数10000以上；头条号指数600以上	功能介绍
热词分析	已开通	粉丝数2000以上；头条号指数600以上；已开通原创权限	功能介绍
扩展链接	已开通	粉丝数10000以上；头条号指数600以上；已开通原创权限	功能介绍
评论保护	已开通	粉丝数10000以上；头条号指数700以上	功能介绍
外图封面	已开通	粉丝数2000以上；头条号指数600以上	功能介绍

图5–9 头条号指数关联的运营工具开通权限示例

自媒体引导

评估，是为了择其优者；服务，是为了悦其善者。客观上看，有人力精细化服务的平台方，一定只会是整个生态的头

部。面对金字塔结构中下部的广大自媒体，平台方会更多地从引导的角度入手，管理好自媒体的预期，让普通作者能够见贤思齐，不断向上前进。

预期管理对所有对接生产端的平台而言（如滴滴、美团、今日头条、微博等）都或多或少地存在。在我看来，归结起来就一句话："生产端能够认知到，通过符合平台规则的持续努力，可以获得一定的预期收益。"

"一定的预期收益"是最终态，生产端通过评估最终态来决定自己投入与否。能够让自媒体保持持续稳定的内容产出，一定会归结到收益层面，这是所有平台都绕不过去的问题。

对个人运作模式而言，"预期收益"是评估依靠平台的分成收入，自媒体是否还算得上是一份不错的第二职业；对工作室或团队运营模式而言，"预期收益"则是对标平台上已有的头部自媒体，来衡量自己借助平台能够成长到多大规模。

"符合平台规则"代表了平台的利益诉求和运营导向。

在不同的发展阶段，平台有不同的发展目标和服务标准要求。无论是奖是惩，平台都应将其落实到规则层面，让运营行为有据可循、有法可依。

相对于扬善的正面引导，惩恶则树立了平台的规则。作弊与反作弊有如魔道斗法，此消彼长。无论发展到哪个阶段，平

台都需要面对自媒体钻空子、违反规则的行为。在这种情况下，统一标准、公开反作弊举措是非常有必要的。自媒体之间也会有沟通和交流，如果看到夺取流量、营收的都是一些伪原创、低质作者，难免会发生“老实人学坏”的情况。

头条号平台关于打击抄袭行为的公告丨2017年9月 *new*	2017-10-09
头条号平台新增PC端手机换绑申诉功能 *new*	2017-09-28
头条号自律组织成立	2017-09-24
头条号平台关于打击刷粉刷量行为的公告	2017-09-14
头条号平台关于打击滥用声明原创功能行为的公告	2017-09-12

图**5–10** 头条号平台规则

对平台而言，规则的透明性与稳定性是一个必须的条件。朝令夕改的平台，只会徒增生产端的不安全感，令其与平台疏离。一个典型的负面例子是：2017年年初，某自媒体平台PR初阶给了自媒体很高的CPM（按广告每千次被展现收费）广告分成，但当发现其预算远超预期的时候，居然在没有通知自媒体的情况下直接砍掉了自媒体的收益。

“持续努力”代表了自媒体的成长性和积累性。

以网约车场景为例，薪酬不倒挂、适度拉开新老司机的差异性和成长梯度，是平台用于管理和引导服务提供者的有效手段。为了吸引新司机而伤害老司机的服务积极性，是一种荒诞的冲刺KPI（绩效评估指标）的短期手段。

互联网平台往往更容易放大消费者的诉求，而忽略服务者的积累诉求，会将每次服务当成独立随机事件。

一个典型的场景：一个用户下单要去机场，用户身边1公里处有一辆新手车，3公里处有一辆老手车，这个单要下给谁？

- 纯粹的效率导向：下给新手车，新手车由于距离因素能够更好地满足用户的需求。
- 实际操作：平台应该做的是在用户可接受的服务体验下，阶梯性地优化有积累的司机的利益。将更好的单子倾斜给老手车，且将这种决策同步给新手车，才能够营造出平台参与者有成长性的群体期望。

对做撮合生意的平台而言，如果不能将生产端彻底自营化（比如神州专车、界面澎湃新闻），那就始终需要解决生产端的预期管理问题。与生态共赢，并不是说说而已。

本书由“ePUBw.COM”整理，ePUBw.COM 提供最新最全的优质电子书下载！！！

常见的推荐问题

在推荐系统的实践中，我们常会遇到各种问题。在这一章里，我们会选择一些常见的问题进行讨论。

推荐重复

在内容生产门槛不断降低、产量持续攀升的背景下，当一个新闻事件发生后，权威新闻源会发布新闻通告，自媒体会从不同角度进行评论，搬运工也会批量产出蹭热点的内容。

内容的大繁荣也带来了信息的过载问题。对用户来说，其需要的是有价值的信息而非千篇一律的内容，用户可能会厌恶列表页上千篇一律的标题，更会因为点击了标题后却发现内容了无新意而怒发冲冠、拍案而起。

重复，对于内容推荐来说，是一个必须解决的问题。

结合用户的使用过程，我们将用户对内容的消费拆解为列表页消费和详情页消费。前者关系到点击前的消费预判（如图6–1左），后者则代表了用户点击后的消费体验（如图6–1

右）。



图6–1 列表页消费和详情页消费

基于这两个维度拆分，我们能够得到一张2×2的表格，如表6–1所示。

表6–1

	列表页相同（相似）	列表页不同
详情页相同（相似）	A	B
详情页不同	C	D

四种场景分别对应了内容重复的四种处理方式：

情况A（甲、乙两篇内容，列表页和详情页都相同）：

从消费角度来看，甲、乙两篇内容对用户来说是具有替代性的。用户消费了甲内容之后，大概率不需要再消费乙内容了。那么，对推荐系统来说，就需要从甲、乙两篇内容中进行选择，选出应当展示给用户的内容。因此，问题从内容选择转变为信息源选择，需要深究重复构成的原因和内容的消费特点。

- 搬运号对原创内容的抄袭：展示原创内容，对搬运者进行惩罚。
- 媒体机构发布的新闻通稿：选择首发、权威度高的媒体，或是基于用户订阅关系，给用户推荐他订阅的自媒体。
- 热点转发内容：在这种场景下“被谁转发”是有一定信息量的。我们会基于用户跟转发者关系的紧密度来判断是否需要展示，紧密度低则不展示。
- 对于某些具有重复消费属性的内容，如音乐、评书、相声等，可以通过产品设计方式（如历史记录、再看一次、收藏列表、播单等形式）让同样的内容重复出现。

情况B（甲、乙两篇内容，列表页不同但详情页相同）：

在转载的情况下，一篇内容以不同的“妆容”展现在列表页当中。这时就需要进一步分析，以确定用户是否看过此内容。

如果用户点击过甲内容，那么给他推荐乙内容的必要性是不大的，因为从点击后的消费体验来看，用户并没有获得额外收益。如果用户没有点击过甲内容，那么乙内容因为发布者、标题、封面的不同，带给用户的列表页消费预判是不一样的，也就有了进一步推荐的必要性。头条号的脑洞功能“双标题＋双封面”就是一个实际的应用。

情况C（甲、乙两篇内容，列表页相同但详情页不同）：

典型的例子如红烧肉的做法、郭德纲最新爆笑相声等，尽管用户点击后的消费体验不同，但是相似的列表页展示会给用户带来消费决策上的困惑。

如果用户点击过了甲内容，他十有八九会以为乙内容与甲内容是重复的，从而忽略；如果用户在列表页看过但是没有点击甲内容，他也会大概率地错过乙内容。对于这种情况，应该拉长两篇内容的推荐间隔，将其视作一个密集打散问题处理。

情况D（甲、乙两篇内容，列表页和详情页都不同）：

这种情况是最简单的。两篇完全不同的内容，互相不构成对用户消费预判的影响，分别推荐即可。

推荐密集

我们承接重复问题继续讨论。密集是指用户的推荐列表中同一类内容的占比过高，导致局部多样性丧失。

导致推荐密集的原因，一方面是因为用户的短期兴趣点通常比较明确，会因为特定事件或人物而快速聚焦。比如，建军节阅兵时，哪怕娇滴滴的女生也会被解放军的飒爽英姿所吸引。围绕阅兵，会产生如军事分析、士兵英姿、爱国抒怀等不同切入点。伴随热点，人们的消费量也会在这一阶段显著攀升。

另一方面则是因为推荐系统对用户的兴趣点理解不够，或是仅追求点击导向而放大了用户的强兴趣相关内容，从而忽视了用户的弱兴趣相关内容。比如，系统只知道用户喜欢财经，就围绕财经内容翻来覆去地推荐，或是在财经内容和科技内容所构成的候选集对比中，由于财经内容点击预估显著高于科技内容，系统就只推荐财经内容。

从点击率角度观测，局部的密集满足了用户需求，往往会带来短期消费量的快速提升。这未必是件坏事。我们在有关用

户冷启动的部分也提到，为了短期留存用户，系统会刻意牺牲多样性，以新、热内容来满足用户的短期需求。

但是，不同用户短期兴趣点的衰减速度是不可预知的。比如，对于阅兵的内容，可能到第三天就无人提及了。有关NBA的赛事报道在赛季结束后，也再无更新。为了防止用户体验发生断崖式的下跌，从产品角度来看，还是希望避免一次刷新中出现内容过度密集的情况。通常，我们采用滑动窗口规则，即连续多条规则尽可能在多个维度打散，降低用户的视觉密集感。

密集的衡量取决于我们对内容的理解拆分维度，能够拆分出的维度越细致，可以做出的打散策略就越细致。典型的可以拆分的维度有题材载体维度、作者维度、类目（话题）维度和实体词维度。

题材载体维度：属于同一内容载体的内容过多，比如一屏幕都是视频或图集。

作者维度：来自同一作者的内容过多，比如高产量的媒体型账号刷屏的现象。

题材和作者属于内容的固有属性，类目和实体词则是基于语义理解抽离出来的属性。

类目（话题）维度：来自同一分类（话题）的内容过多，

比如阅兵期间有关阅兵的内容大量展现。

实体词维度：讨论某一实体词的内容过多，比如都是关于某位明星名人的各种维度的论述。典型的例子如马云，横跨财经、科技、教育等多个领域，在系统对于类目识别准确的前提下，依靠类目打散都不能解决马云刷屏的问题，只能通过实体词进行打散。

关于密集打散的收益，我们通常会站在用户长期留存的角度去衡量。多样性更好的内容在短期可能会降低点击率，但长期来看，对用户留存是有帮助的，不必争朝夕之短长。

易反感内容

在内容推荐系统中，我们将用户的行为拆分为列表页的消费决策和详情页的消费体验两部分。无论是详情页的消费体验差还是列表页的消费体验差，都是需要处理的问题。如图6–2所示，列表页中展现出了标题和封面信息供用户判断，用户点击了前两篇内容，没有点击第三篇内容。

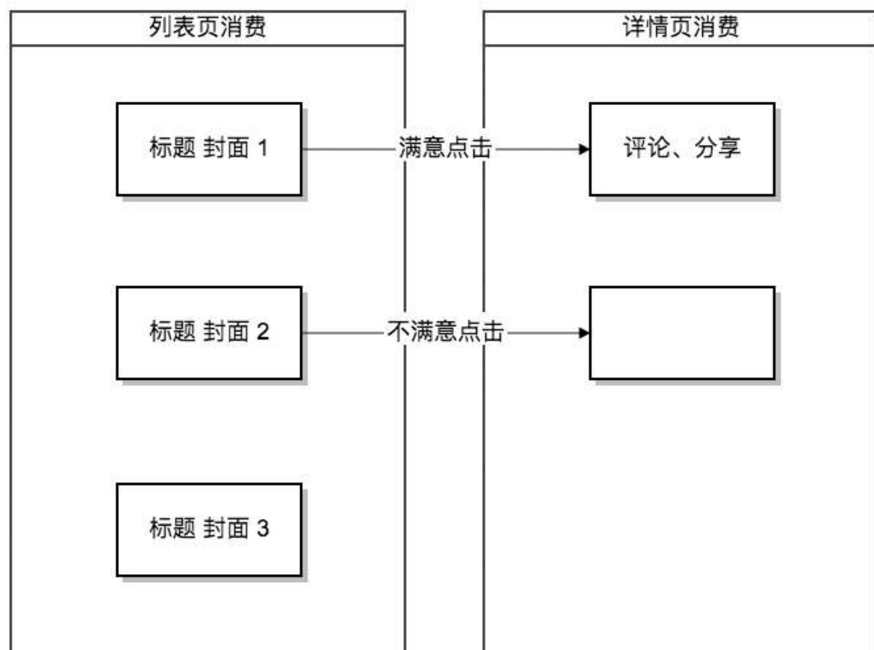


图6–2 列表页消费与详情页消费

从内容质量维度看，低质量的内容一定会引发用户的反感，如文不对题的标题党、传播虚假信息或耸人听闻信息的猎奇党、质量低下的无聊水文、因时间识别错误导致的旧闻问题等。如图6–2中第二篇内容的情况，用户在被标题吸引点击后可能很快就会退出，或是在内容当中举报，留下负面评论内容等。我们可以通过用户的反馈行为来发现第二类内容，以降低它们对用户体验的影响。

但是，对于列表页刷过但是没有被点击的内容（如图6–2中第三篇内容）又该怎样处理呢？

通常，我们将没有被点击的内容视作对用户无损，它起到

了如兴趣探索、广告变现等作用。但在实践过程中，我们发现有部分内容会因题材问题而非质量问题，对部分敏感用户的列表页消费体验造成负向影响，我们将之归结为易反感内容。常见的易反感内容有：

- 鬼神类内容：如灵异故事、UFO（不明飞行物）等
- 恐怖血腥类内容：如蛇、野生动物等
- 迷信类内容：如算命、风水等

对易反感内容而言，先要识别得准，再要推荐得好。

识别过程采用的是数据产品中定量分析处理问题的通用流程：分析影响面、定义标准、数据收集标注、模型评估。



图**6–3** 易反感内容示例

分析影响面旨在帮助我们更好地确定待解决问题的优先级，将有限的精力优先投入到收益场景更大的事情上。就易反感内容的影响面而言，用户的反馈量统计就是一个衡量标准。将一定周期内用户的反馈进行整理和标注，对应到不同的易反

感内容类别上，就得到了各个类别的影响面情况（如表6–2所示）。

表6–2 用户反馈量统计示例

易反感内容类别	反馈量	处理优先级
鬼神类内容	10 例	中
恐怖血腥类内容	53 例	高
迷信类内容	5 例	低

在确认了问题的影响面和优先级之后，就需要针对各类问题达成一致可理解的定义。在这一环节中，我们需要抽离出问题最重要的抽象信息，让第一次看到这些信息的人也能迅速明白这类问题的主要表征是怎样的；然后，辅以大量的正例、负例，以帮助参与的人能够通过例证更好地对齐理解、把握尺度。比如，对迷信类内容的标注为：

表6–3 迷信类内容标注示例

说明	样例内容
与身体部位有关的迷信类内容	• 拥有这手相的人，难有作为，可谓劳碌命 • 拥有这手相的人要警惕，中年横灾或大病 • 这些部位显示出女人婚姻不幸福 • 这三个地方长痣的女子，一生操劳婚姻不顺 • 最近身上有这 3 种变化，预示必有横财到来、富贵盈门
风水禁忌类内容	• 卧室有这几处风水，就是给财神爷穿小鞋，运势下降，霉运也缠着你 • 让你一辈子都一败涂地的风水 • 属虎的人不得不知的 5 个风水禁忌
诅咒类	• 命中注定“克夫”又“克父”，婚姻不顺的生肖 • 这个生肖的人今年婚姻出现危机 • 婚姻不顺，今年会有离婚信息的生肖女 • 属兔、属蛇、属龙人在几岁的时候最倒霉，厄运频繁，小人中伤 • 厨房染上这些恶风水，霉运蔓延，财运溃散，谁都帮不了你

数据收集与标注是为模型训练做准备的。基于对标准的理解，标注人员通过双盲校验的方式标注出足够进行模型训练的数据。在大型互联网公司，通常都有专门的数据标注人员负责企业数据的标注和整理。

在研发产出识别模型后，产品经理还需要对模型的准确度和召回度进行评估。在不同的应用场景下，准确和召回的平衡度是不一样的。在易反感问题上：如果要强化读者端的体验，尽量不让用户看到易反感的内容，就需要重视召回率，宁肯误删也不能漏删；如果要强化作者端体验，希望作者的内容尽量不要被误删，则需要重视准确率，适当露出一些置信度低的结果或增加人工复审捞回流程以保证内容不至于被错杀。

在推荐过程中，对易反感内容的推荐处理是一个强化负反馈的过程。

在列表页展示层，对普通内容而言，如果用户不点击的话可以视作无损，而对于易反感内容，即使用户不点击，也应当视作一个负向反馈，拉长此类内容的二次探索周期。

以推荐蛇和NBA内容给女性用户为例。如果用户都没有点击，那么可以将NBA内容在下一个周期（比如5天）再次推荐给用户，作为二次探索尝试；但蛇的内容就应该推迟多个周期后才推荐，以降低对用户可能的影响。

在令用户反感的行为上亦然。用户往往只会点击页面上的关闭按纽，而不会选择具体原因。如果一篇内容同时命中多个点的话，容易引起反感的原因理应受到更大的惩罚值。

由于易反感内容多发生在探索环节，主要影响用户的列表页体验，故在统计上并不会显著影响留存。但对此类内容进行更细致化的处理印证是做内容分发的初心：如你所愿，阅你所悦。

时空限定内容

作为一种消费品，内容是有特定消费时空限定的。如果以商品做对比的话，就如同端午节卖月饼、在上海卖豆汁儿一样，销量必然惨淡。深入理解内容的时效性和空间性消费特点，能够帮助我们更好地提升消费体验、促进内容消费规模增长。

在时效性上，不同的内容有不同的保鲜期。

短时效性内容，以赛事、股市信息为代表。这类内容具有保质期短、时效性强的特点。它们可能会每半小时更新一次。在新版本的消息出来后，旧的消息就完全没有价值了。比如：演唱会预告的内容，就不应该在演唱会之后再推荐出来；赛事结束后，中场的消息就不应该再展示了。

中时效性内容，通常可以覆盖绝大多数新闻内容，涉及最近发生的新闻事件，以天和周为时效性周期。比如，与电视剧《权力游戏》相关的内容生产与消费通常与电视剧播出节奏保持一致，在剧集播出之后，其热度逐渐衰减。

长时效性内容，具有跨时间维度的消费价值，可以是知识、案例分析等非虚构内容，也可以是小说、散文等虚构内容。

对大型推荐系统来说，在处理内容时效性问题上，需要平衡两方面的问题：长时效性内容推荐的日期短了，会造成资源浪费；短时效性内容推荐的时间长了，会对用户体验造成伤害。因此，推荐系统会基于内容的特点预判不同内容的衰减周期和推荐策略。

对于短时效性内容，在识别层面，可以首先基于消费规模进行头部类目的覆盖（如介入可枚举的天气、股市、赛事、时事等信息），其次进行通用的时效性识别。在推荐层面，短时效性内容需要更快速地传播到用户消费列表和推送场景上去。如果一篇内容正常需要一个小时的缓慢冷启动过程的话，短时效性内容可以依赖规则触发机制，在半小时内完成全量传播。比如天气、股市等信息，是完全可以对接权威官方机构，实现自动化分发乃至推送的。

对于中时效性内容，大多数在两三天的维度内均可消费

（如娱乐、汽车、游戏、行业分析等），系统通常会视自己的内容产出速度来制定对应的衰减周期。大平台的衰减周期短一些，小平台的衰减周期长一些。

对于长时效性内容，最典型的应用场景就是搜索，什么时候被用户主动检索出来都具有一定的消费价值。对推荐系统而言，越大的候选集理论上能够产出的推荐效果越好。所以，我们实际做的是计算资源与点击收益的平衡，即什么样的内容值得被长时间保留在推荐候选集当中。用内容在垂类中的消费量衡量是个相对简便的方法。比如，围棋视频的平均消费量是1万次，而其中柯洁大战AlphaGo（阿尔法狗）的消费量是10万次，那么，就可以把它视作围棋领域内相对经典的内容，从而获得更长的推荐周期。

除了时效性，内容还具有地域消费性的特点：即时天气、本地新闻、同城活动、新店试吃等内容，都应当局限在特定的城市进行分发。如果分发到错误的城市，内容的实用性就会降低。

对于地域的识别，我们通常从正反两方面着手。正面的地域识别，可以通过标注本地媒体的方式来区分内容分发的范围（如《金陵晚报》的内容直接往南京分发），也可以通过内容中的关键词密度进行补充识别，如大篇幅提到海淀、朝阳，可能就是北京地区的内容。反面的识别豁免，需要对如旅游、历

史等类目的内容进行豁免，此类内容中会频繁出现地名，容易造成误伤。

本书由“ePUBw.COM”整理，ePUBw.COM 提供最新最全的优质电子书下载！！！

带着偏见看推荐

也许，这一章的名字叫作“竟然被骗了？！关于算法推荐的3个误解，快来看看你有没有？”效果更好。

推荐其实不是一个新事物，它早已被应用在了如淘宝购物推荐、豆瓣电影推荐、点评美食推荐、携程差旅推荐等不同的场景中，以提升服务质量、优化系统效率。

当推荐技术应用于内容分发领域后，在已有传统媒体、门户网站仍然占据重要分发位置的情况下，质疑的声浪频传。如果你经历过PC时代，也许会觉得这一幕似曾相识：当年，传统媒体对门户发出过质疑声，中心化的门户模式对去中心化的微博模式同样质疑过。只是岁月轮回，昨日的挑战者成了今日的卫道士，何其唏嘘。

在本章里，我所做的并非辩护，因为真实的数据表现和传统新闻应用的推荐化转型都充分证明了内容推荐模式是有一定存在价值和发展空间的。我试图从自己的认知角度，对一些常见的误解进行阐述供大家参考。

站在从业者的角度，我觉得现在的内容推荐产品还远远没有达到理想态，其迭代过程中是存在种种瑕疵和待改进空间

的。但发展的进程并不会因为人们的误解而止步。我坚信，科技产品的存在是为了让生活更加美好。

信息茧房

当内容分发全面进入推荐分发时代，对这一分发模式的质疑也就越来越多，很多人站出来号召大家要一起戳破“算法的泡泡”。

一种普遍观点是：机器算法主导的精准分发，在提升阅读体验的同时，也极易导致由信息收窄带来的一叶障目。即在算法分发下，用户将深陷信息茧房当中。

哈佛大学法学院教授、美国前总统奥巴马的法律顾问凯斯·桑斯坦在其2006年的著作《信息乌托邦——众人如何生产知识》中提出了“信息茧房”这一概念。通过对互联网的考察，桑斯坦指出，在信息传播中，由于公众自身的信息需求并非全方位的，公众只注意自己选择的东西和使自己愉悦的领域，久而久之，会将自身桎梏于像蚕茧一般的“茧房”中。

这顶“信息茧房”的帽子，随着信息流量分发的迁移，被扣到了不同的内容分发服务头上：

在国外，先后被吐槽的是谷歌的个性化搜索结果和脸谱网

的信息流服务。

在国内，早在2012年就有人吐槽微博是“信息茧房”，如今吐槽的对象又变成了机器推荐分发。

当信息生产的门槛不断降低造成内容量的大繁荣，当信息消费者的选择权越来越大、越来越能够主动选择而非被动接受的时候，令人担忧的“信息茧房”就会被越来越多地提及。

然而，过滤你的并不是算法泡泡。

在纸媒时代，当用户从特定的媒体人、特定的媒体刊物处获取信息的时候，其信息获取方式不就是纸媒版的订阅关系分发吗？各家纸媒有自己的题材偏好和内容风格，不也构成了一个“茧房”吗？当面对报亭中琳琅满目的刊物和邮局的订阅表时，用户的主动选择便构成了他的认知世界。

还记得那本月发行量逾700万册的杂志——《知音》吗？其刊载的内容正是今天饱受诟病的典型标题党样例：《风之谷啊我的妹妹，哥哥的未来献给你》（2007年第7期）、《再大的恨放下吧，唤醒前夫赢得亲情一片天》（2007年第35期）。

在10年前，为什么会有那么多人消费这本杂志，在10年后

的今天，就为什么会有同样规模的人在消费着类似内容。从用户需求的角度看，这些内容符合大众用户偏好、可以满足他们打发时间的需求，高阅读量就是用户主动选择的结果。

内容作为一种消费品，每个人都有选择消费途径和消费内容的权利。无论是纸媒还是网媒，只要有足够大的候选集和主动选择权，用户就一定会选择自己更为偏好的信息载体和信息源。从内容匮乏到内容繁荣，从中心化一统到垂直化聚群，用户的选择更贴近自身喜好是不可逆转的趋势。

那些持续表示担心的人，本身就是多源信息的消费者，希望看到多源信息的渴望胜过了他们看到不感兴趣内容的不悦。而被担心的人，本身就不曾有信息多样性的焦虑，乐得在自己的小圈圈里打转。这就形成了一边有人大声疾呼，一边有人甘之如饴的局面：担心的人不曾被过滤，被过滤的人不曾担心。

进一步，对那些担心的人来说：他们跳得出算法分发，也跳不出社交关系分发；跳得出社交关系分发，也跳不出自己的认知选择边界。“茧房”始终存在，真正过滤你的，是你的认知泡泡。

在这种情况下，将“信息茧房”归罪于分发模式更像是因失语而产生的牢骚。与其吐槽分发模式，不如考虑在新的模式下

如何优化；与其大声疾呼内容分发服务商的道德义务，不如将其归结为内容服务商的利益诉求。

要知道，“信息茧房”并不是内容服务商的诉求，“用户茧房”才是其诉求。为保证用户更好地长期留存，内容服务商就势必需要不断探索用户的内容兴趣偏好、了解用户的观点偏好：

你以为只给用户展示他喜欢的类目就够了？那你错过的就是拓展用户兴趣认知的机会。只有给用户展示多元化的内容，才能够让他消费更多内容，停留更长时间。

你以为只给用户推荐他偏好的观点就够了？那你就可能错过用户间的讨论和互动碰撞。给用户展示多元化甚至有冲突的观点，对评论的刺激作用是显著的。而在多元化的讨论之后，内容丰富的评论区也具有了可以消费的价值。

如果你的认知还停留在算法分发只会给你推荐感兴趣的内容上，那你的认知泡泡，也该戳一戳了。

推荐会导致Low？

对内容推荐的又一个常见误解是：推荐系统会趋向于低质量和Low（低格调）的内容。做出直观判断很容易，深究这一

判断背后的原委却并非易事。那么，让我们来仔细讨论一下。

1. 三问内容质量

既然说内容质量低，那总得有个标准吧，这就引入了第一个问题：什么样的内容是质量好的内容？

当我向朋友提出这个问题时，大家往往会陷入短暂的沉默：是啊，对于食品，我们有诸如保质期、配料成分表、制作工艺等方面的要求；对于衣物，我们有型号、材质、洗涤方式的规范；对于内容，这种非标准化的手工产品，是否也存在一套类似ISO9001的标准可以衡量呢？

幸亏身边有纸媒的前辈，让我得以请教纸媒对内容的衡量标准。其曾供职的报社，对优质内容的衡量标准如下：

- 受众角度：读者关注面广，在社会上引起较大反响，为报纸争得明显的效益或荣誉，收获普遍好评的热点、焦点新闻稿。
- 策划角度：富有策划、创新意识，极具冲击力和感染力的报道。
- 深度角度：分析透彻，有独到见解的深度报道。
- 题材角度：重大独家新闻，重大调查性报道。

“较大反响”“深度”“重大”“普遍好评”，我们不难发现，上述标准多为难以量化的描述性用语。除了第一条站在受众角度——“读者关注面广”“较大反响”，我们尚可以用阅读量、转发量、评论量等数据指标度量外，第二条到第四条都是从创作角度出发，属于主观判断的范畴。对特定垂直行业来说，资深从业者或许有资格能够进行内容深度的判断。对平台来说，通常会涉及多个垂直行业，我们只能以频道编辑建制的方式来解决这个问题。

大众说好，专家也说好，“叫好叫座”大概是好内容最理想的终局，一如触乐网的那篇《在三和玩游戏的人们》。

在达成“好”的共识之后，进一步的问题是：

如果一篇内容专家不叫好，这样的内容是否应该被过滤掉呢？

未必。

在看过爆米花电影之后，我有翻豆瓣电影看影评的习惯。可看完影评之后我常常会自惭形秽：明明自己觉得还不错的电影，为什么在影评人眼里变得满目疮痍，充满各种各样的问题呢？自己觉得还不错的桥段，在影评人眼中却成了陈词滥调，缺乏创新与深度。

在猫眼电影工作的朋友的一番解释让我释然：“你是观众视角，影评人是专业视角。专业的判断跟大众的喜好通常会存在认知背景的偏差，在技法上有待改进的内容并不意味着缺乏受众。一如范雨素的爆红，从文字技法上看，她的内容一定是有缺失的，但她对生活的记录触动了许多人的心。”

诚然，如果交由特定群体来判断，或许永远无法产出如“快手”这样的产品。同一段小视频，你看到的是残酷底层物语，他看到的则是生活的延展。

在今日头条上，同样有这样的账号：农民王小。用直白的影像记录生活，记录老妈做饭，记录姐姐、姐夫给爸妈送米面油……视频的艺术性？连我这个外行都觉得谈不上。但就是这样的账号，已经累积了45万名粉丝，每一条视频都有几十万甚至上百万次的播放量，上千条评论。



图7–1 今日头条“农民王小”账号内容

这样的内容，我们怎么评价？它也许不能称之为“好”，但也绝对不至于被定位为低质的内容。正如面对鸡蛋与石头的冲突，我们或许首先应该倾向于鸡蛋。如果进行过度专业化的过滤，或许这样的内容将永无出头之日。

那就回归到最后一个问题：什么样的内容是低质的，应该被过滤？

虽然对内容质量的判断存在主观性，但并不意味着我们丧

失了对内容质量的约束。借助大量的案例分析，我们能够抽离出一些客观指标，也能达成一些平台审核层的主观一致标准，以较少争议、控制误伤量的方式给出评判“内容质量差、不宜传播”的标准：

- 真实性上：歪曲事实，虚假信息等。
- 阅读体验上：无意义内容，错别字连篇，文不对题，语句不通顺、掺杂广告等。
- 价值引导上：不正确的价值观导向，煽动对立情绪，低俗色情等。

我猜想这也是快手对特定内容做出干预的原因，记录这些内容也许无罪，但这些内容并不适合传播。在明确的低质内容底线之上，平台方或许应该对内容和内容创作者保持足够的宽容。

2. 为什么会觉得推荐内容格调低劣

既然推荐系统已经将绝对低质的内容过滤掉了，那么，为什么我们还会时不时地觉得推荐内容格调低劣呢？这主要有以下两个原因：

其一，推荐准确性问题。

一方面，每个人的认知程度是不同的，如果一篇内容是低于我们认知水平的，那我们一定会觉得它质量一般、内容格调偏低劣。就像知乎的Live产品，不管是受到多少好评的讲座，总有人会给出中评或差评：“格调太低劣了，没啥干货。”另一方面，我们对某些内容存在个人偏见。比如，有些人认为郭德纲的相声是“三俗”的，有些人觉得八卦新闻是不好的，等等。如果推荐不能很好地匹配用户的兴趣偏好或认知程度，就会让用户产生内容格调低劣的感觉。

在冷启动阶段，由于对用户缺乏认知，推荐并不能很好地刻画用户的兴趣画像。为了最大化地实现用户冷启动阶段的留存目标，推荐必然会从大众化和基础化的热门角度切入。而大众化的内容必然是与部分用户的认知水平存在差异性的，从而让这部分用户认为推荐内容是低质的。

随着用户的点击动作、阅读表现、关注行为的积累，基于用户聚类和用户偏好作者的信息，推荐不准的情况能够在一定程度上得到缓解。比如，一个人常看科技内容，不点击娱乐内容或很少看完八卦新闻，系统会大体上知道他的调性，不再推荐他不感兴趣、觉得格调低劣的娱乐八卦新闻。

在推荐过程中，更难解决的是认知水平的问题。系统基于文本分析确定了内容分类，可能会推荐一篇质量平平的科技分

析给一个重度的科技用户。在这种情况下，用户还是会觉得内容过水，从而对推荐产生微词。我们通常依赖用户对作者的关注程度来局部改善此类问题。

其二，用户无意识反馈问题。

先讲个真实的案例：

一日聚餐，席间A君说："你们这个不行啊，老是给我推娱乐八卦，应该多给我推荐一些行业观点、深度分析之类的内容。"

B君道："好好好，把你的用户ID给我，我查询一下，看看推荐算法是不是有什么问题。"

半晌，剧情反转。

B君朗朗道："下午1点半，刷新一次，给你推荐了3条行业资讯、2条体育资讯、1条娱乐八卦，你点了娱乐八卦。下午2点半，刷新三次，你点击了1条行业资讯、3条娱乐八卦。下午3点，刷新……"

众人相视爆笑，终结谈话，定论："本我超越了超我。"

想象一下，当一篇八卦新闻和一篇深度分析同时出现，你会做出怎样的选择？站在马斯洛需求模型的金字塔前，本我制造了足够大的需求。你点击一篇娱乐八卦，很有可能是本能驱动的下意识行为。你更多地点击娱乐八卦又给了推荐系统对于此类内容更强的反馈信息，从而增加对此类内容的推荐。如果从推荐系统的点击预估角度看，更接地气的内容超过高大上的内容几乎是必然的。

消费八卦新闻的确帮助你打发了时间、给你制造了欢愉，但你偶尔的“超我”自觉自省，却对这种内容推荐做出了格调低劣的判断。这也是我们在用户访谈过程当中常会碰到的一种情况：就像在一个个夜里，你一遍遍地跟自己说要早睡，却一次次刷着王者荣耀直到深夜。

3. 更平衡的产品设计

尽管推荐技术存在一定局限，尽管用户反馈总是会表现出泛娱乐化的趋势，但一个理想态的产品是不应该仅以点击为导向、唯短期数据而论的，我们能够做出的产品设计不应该止步于此。

假如我们在点击一篇内容重复度很高的娱乐八卦和点击一篇行业深度分析之间，倾向于认为后者更有价值，那就需要找到更好的方式来衡量一次阅读的价值，从而引导推荐系统的分发流程，比如：

- 内容稀缺度：越垂直的内容越稀缺，小类目下的内容点击可能比大类目下的内容点击更有价值。

- 作者角度：从全局来看，每个垂直赛道都会跑出有广泛知名度的内容品牌，它们的内容往往更适合广泛传播；从用户个体来看，某些用户会对内容来源而非内容类目更加敏感，这就需要放大作者的相似度以降低内容的相似度。

- 互动行为：不同的行为代表了用户不同的意图。阅读行为仅仅是一种个体行为，点赞、评论都代表了用户对内容更感兴趣，而分享则意味着用户愿意为之传播和背书，可能权重更大。

不同角度的分析给我们提供了量化点击价值的方式，让那些我们更偏好的内容有可能获得更多的曝光量和展示量。从现实的角度来看，内容供给和内容消费一定会是金字塔结构的，越基础层的内容越具有消费规模。但作为理想的现实主义者，我们或许可以在可控的范围内，去尝试点击率和理想态的平衡。

编辑、算法与社交，三分天下？

先抛出个人的结论：编辑（中心人工主导）分发、算法

（机器主导）分发、社交（离散人工主导）分发各有千秋。内容分发服务追求的是分发所能触及的这一远景，为了达成这一远景，就需要探寻每一种分发更适合的应用场景，而不是要在“剑宗”和“气宗”间争个高下。

本节试图探讨的，就是每种分发方式适用的场景和它们在内容分发系统里的应用。

编辑分发vs算法分发

从纸质报纸、杂志到广播电视再到门户网站，尽管信息传播的载体发生了变化，但是内容传播始终保持着“中心化分发，展示位有限、千人一面”的状态，信息传播的决策权始终握在编辑手中。

编辑分发的优势在于，借助专业背景知识完成了从海量内容到有限展示位置的过滤和筛选。经过筛选的内容，其平均质量是相对较高的。然而，基于专家判断的分发难免会出现偏差，为了降低“叫好不叫座”或单个编辑偏差的情况，内容分发方也采取了相应的策略。比如：传统纸媒会有编委会投票机制，通过多人判断选题；门户网站有点击率退场机制，点击率在一定时间内不达标自动下线等。

引入了机器推荐算法的分发系统，由于达到了“千人千面”的效果，展示位数量得到了大量的扩展。在筛选人力不足

以匹配展示位数量的情况下，编辑又起到怎样的作用呢？

首先，人工同机器一定不是对立的，不然，今日头条也不会吸引如此多的资深传媒背景从业者。我和组内的同事不止一次地真心赞美吴达（时任头条号运营总监）：“我们的运营团队都是豪华高配的文化人。我跟吴达老师聊天是要带字典的。”

在内容层面，编辑和审核团队能够决定什么样的内容是低质的，不应被系统收录和推荐。编辑和审核评估团队就像是内容推荐系统的门神一样，对不好的内容可以说不，对低质内容背后的做号者也可以说不。

作为最大的信息和社交分发平台，脸谱网也在事实甄别和低质内容管理上强依赖编辑和审核团队的人工工作。

2016年12月16日，脸谱网上线了Fact-Checking（事实审核）机制，将用户举报过多的信息交付机构记者来判断。如果记者判断这则内容是假新闻，该机制就会将这一内容标记为存在争议，一方面会在前端页面提示用户此内容可能失实，另一方面会从分发量的角度进行控制。

2017年5月，扎克伯格发帖称脸谱网会再招聘3000名内容审查员，在此次招聘后内容审查员将会达到7500人。审查员会过滤社交媒体上的不适当内容，如恋童癖、身体暴露、种族仇

恨等内容。

除了说“不”之外，编辑同样会对特别值得推荐的内容说“是”。

以推送场景为例，作为强打断的场景，被推送的内容需要被审慎地筛选出来，以保证打断用户是值得的。新闻客户端一直强调筛选出“Breaking News”（突发新闻），当重点新闻发生时，编辑一定是24小时值守的，并验证其真实性，以确定推送范围和推送级别。只为了不错过每一条值得用户关注的内容，让用户能够更准确快速地获得最新消息。在这一过程中，技术能够辅助编辑更快速地构建新闻候选集，比如追踪社交媒体、重点网站的发布情况（如气象局、交通局）等，但人工才是担任最终裁决的角色。

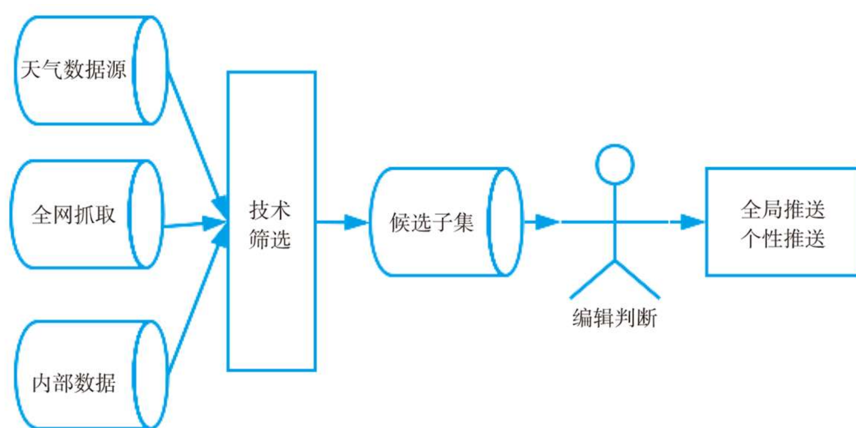


图7–2 推送候选集的筛选逻辑

当然，为了保证系统的可扩展性和有效性，我们希望在日

常推荐中尽量避免人工的直接干预，如对内容进行调权、对展示量进行干预等。但是，不直接干预并不代表缺位，编辑始终在扮演“纠偏”的角色，当发现主观上觉得好的内容没有得到应有推荐量的时候，当发现主观上觉得差的内容得到过高推荐量的时候，都会给产品和技术做出反馈。在这种情况下，产品、技术、编辑（内容运营）会坐下来探讨：第一，这是不是一个问题，如果是问题的话，是不是一个频发的问题，以此来确定解决与否和解决的优先级；第二，探讨更系统的解决方案，让这一类而不是这一个问题得到更系统的解决。

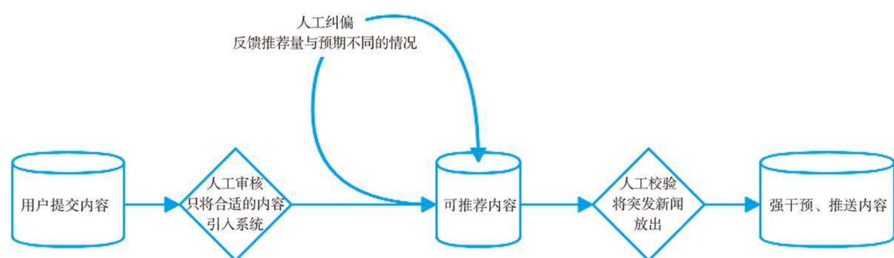


图7–3 人工在内容分发流程中的参与角色

编辑不仅能够帮助推荐系统更好地理解内容，也能帮助我们更好地理解站在内容背后的创作者群体。

在创作者体系层面，我们常说：做产品要有用户视角，作为平台方，你在面对亿万用户的同时，也在面对万量级的作者。作者同样是内容分发系统的用户。编辑会更理解创作者的语境，成为创作者和作品的代言人，从而影响系统的迭代方向。与此同时，他们也能够以创作者更易理解的方式去传递平

台的规则，帮助不同阶段的创作者在平台更好地成长。

社交分发vs算法分发

在脸谱网、微博、微信覆盖了越来越多的用户之后，内容的分发逐步去中心化：每个人都可以创作内容从而成长为自媒体，每个人都可以借助社交关系评论、转发从而完成内容的传播。信息的传播权从传统的精英编辑让渡到每个普通受众，相当于每个人都成了编辑，成了内容分发的中心。

社交分发首次让信息传播变成“千人千面”。每个用户都有了个性化的内容消费。2010年，脸谱网主页访问量超过谷歌访问量，意味着社交分发已经成了主流的分发方式。援引皮尤研究中心此前的调查，美国成年人中有62%通过社交媒体获取新闻，有18%高度依赖该平台，通过脸谱网阅读新闻的人数占比高达44%。

当然，社交分发也带来了新的问题：

一方面，进入稳定期后，流量出现了新的垄断：一些大号由于拥有大量的粉丝、保持了高频的发布量，实际掌握了平台的流量分配权。比如，微博上大部分流量一度被营销号和大号所占据，使新的内容生产者获取流量的成本剧增。

另一方面，随着社交关系的不断扩张，微博、微信已经成了线上名片，用户关注了越来越多的来源，基于社交分发的内

容质量也逐步下降。朋友圈中盛行的养生文、微商广告、晒娃帖等就是最好的例证。

社交分发在让人们免于信息匮乏的同时，也带来了信息过载的问题。为了优化用户的信息消费体验，脸谱网率先在自己的News Feed中应用算法进行排序。

脸谱网最初的排序方式，称为边际排名算法（Edge Rank Algorithm）。该算法的核心计算公式为：$E = u \times w \times d$。

- u：用户与内容发布者之间的亲密度分数，互动越高的亲密度分数越高。
- w：不同反馈动作具有不同的权重，反馈动作包括展示、评论、点赞等。比如，评论动作的权重就会显著高于点赞动作。
- d：基于时间的衰减，越新的内容权重越高。

通过上面的公式不难看出，亲密度和用户动作的引入，极大地抑制了大号和营销号刷屏的情况。此前，企业账号一旦获得了粉丝就相当于获得了稳定的广告位，所有新广告以几乎零成本的形式展现在这些粉丝的信息流中。但此后，没有互动的粉丝就只是停留在页面上的一个数字而已，企业账号在获取粉丝之后，必须同时下力气来维护自己的粉丝群体。

在随后的日子里，脸谱网致力于借助机器学习的方式改进排序算法，除了最初的3个Edge Rank因素之外，不断追加新的特征和排序方式，如Story　Bumping（系统对用户错过的信息进行二次判断，如果判断为重要，则会跳过时间序进行置顶展示）、Last　Actor（系统根据用户最近频繁互动的50人，进行信息排序的调权，放大短期兴趣的影响）等。



图7–4 脸谱网改进排序算法示例

无独有偶，在国内，微博也逐步意识到自己的流量被大号和营销号所挟持的问题，开始越来越多地在信息流上应用推荐算法，将原有的时间排序调整为智能排序，以控制刷屏、广告泛滥等问题，优化用户的使用体验。

微信生态在创立伊始就刻意避免微博曾经碰到的问题，更刻意避免流量集中的问题。如果你关注过微信朋友圈的分发就会注意到，某些疑似过度传播的内容（微商广告、转发集赞）是会被微信降频过滤的。

微博、脸谱网将关注关系作为筛选因素，将用户的点击、评论行为作为调权因素，是在“关注关系产出内容”的候选集上进行算法排序。

相比起来，今日头条将关注关系也弱化为调权因素，从而获得了一个更广泛的候选集范围（相当于用户在今日头条上关注了所有头条号），在此之上进行的排序能够有更高的效率匹配性和更好的可扩展性。

关注对推荐系统而言，既是场景，又是助力。

内容推荐是一个预期不稳定的场景，用户持续地刷新、阅读内容，其内容可能是娱乐、体育、社会新闻、财经报道等，不一而足。关注场景给了用户一个稳定预期消费的场景。用户知道这里的内容是产自一个限定候选集合（微博的订阅号列表）时，甚至会主动找寻特定发布者的内容（微信的订阅号列表）。

有效的关注依赖用户的自知自觉，即明确关注是怎么一回事，知道关注后去哪里消费，关注频道会变成什么样。如果不

经平台的干预，很有可能会重蹈社交分发平台的覆辙，大量无效的关注关系会影响用户的关注信息流。

关注对内容推荐还有助力作用，内容推荐系统将关注动作视为用户表意性更强的动作。它代表了用户对某个品牌的足够认知。在这种品牌认知下，用户可能会放松对内容表现形式和内容消费价值的要求。比如，当你关注一个账号的时候，你可能会更愿意消费作者分享的私人信息，阅读那些标题封面看起来“素面朝天”的内容。

关注动作也能被内容推荐系统用来衡量作者的价值，粉丝的多寡、粉丝阅读率的高低都是作者表现的量化指标，一个垂直的、能够持续取悦粉丝群体的作者，才是系统内价值更高的作者。一种应用方式是：某订阅分发平台对某条内容的推荐会先尝试10%的粉丝，如果粉丝点击率高的话会再继续扩散，否则会停止内容的继续推荐。

算法分发：分发的终局？

从某种角度来看，算法分发或许可以被称为终极解决方案。

为什么这么说呢？因为推荐算法是个筐，什么都能往里装。算法是基于我们对现实世界的理解进行的抽象和建模，所有我们关心的因素（编辑分发、社交分发）都可以转化为算法

推荐的参考因素。

如果我们简化算法推荐过程，将推荐的因素收敛到编辑、社交、模型三种因素，那么，一个内容在系统中的得分可以表示为下列公式：

内容得分＝ a×编辑因素 ＋ b×社交因素 ＋ c×模型因素

a、b、c分别为三种因素的权重。如果我们把某种因素的权重设置为1，其他因素的权重设置为0，那么算法分发就能够等同于编辑分发或社交分发。

在这个公式中，各种权重的调节完全是由平台的价值感导向决定的。以脸谱网为例，其认为来自真实好友关系的生活记录内容更重要，在分发过程中就会加强真实好友生活记录内容的权重，而弱化他们转发内容的权重，并进一步弱化媒体所发布内容的权重等。

事实上，我们熟悉的各类内容分发产品，无论起步如何，如今都走上了一条多元素融合的道路：依赖中心化编辑引导和干预，依赖去中心化用户生产传播，应用机器学习提升效率。新版的微博在关注频道的旁边放置了热门频道，提供不依赖订阅关系的内容推荐服务；微信亦上线了实验室功能“看一看”，

尝试推荐分发。

多分发方式的融合一定是未来分发的主流，沉迷“气宗”“剑宗”之争的看客们，还是散了吧。

本书由“ePUBw.COM”整理，ePUBw.COM 提供最新最全的优质电子书下载！！！

面对推荐系统

从茫然无知到过度自信，再从过度自信到心存敬畏，对于自媒体和内容分发业务，我的感受如是。

还记得早年间，《人人都是产品经理》一夜间将“用户视角”这一概念进行了科普。那时，产品经理最常挂在嘴边的话是：“站在用户的角度，我觉得某功能可以上线。”拍脑袋上线的功能比比皆是。

而随着数据驱动迭代流程的日益完善，千人千面的策略成为现实时，产品经理的措辞变为了：“基于统计数据，我觉得某功能可以进行AB实验。”

从主观到客观，从自称典型用户到承认自己只是一名非典型用户，我们也许需要从第三方视角，基于统计数据去观察揣摩我们的用户。

人机大战：效率与目标之争

输给机器不冤

1997年，深蓝战胜了卡斯帕罗夫。

2017年，Alphago战胜了柯洁。

在我初入百度时，系统里还有很多陈年的人工干预规则。

而在此刻行文时，百度公布2017年第二季度财报，其信息流业务DAU（日活跃用户数量）破亿元，全面转向人工智能战略。

作为有技术背景的产品经理，我坚信：在给定的框架和目标下，人类可枚举的规则是远逊于机器学习得来的模型的。

举个例子，如图8–1所示：

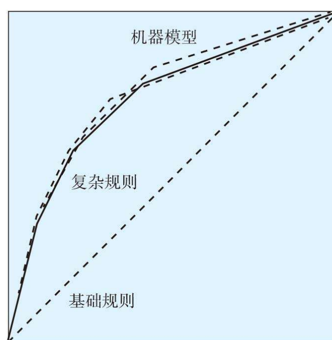


图**8–1** 人为枚举规则与机器学习模型的复杂度对比

产品经理最直观能够想到的策略如图8–1中直线，简单但覆盖面积小。为了提升覆盖面，给策略增加了许多分支判断，从而构成了图中的折线。完成了以上两步，人脑能够实现的复杂度基本到顶了。想要拿剩下的收益，只能依靠机器。图中的

曲虚线可以视作机器习得的模型，通过更高的复杂度覆盖更多的精细场景。

援引戴文渊，即“戴神”的一段经历：“2009年，我加入百度的时候，所有搜索、广告都是专家规则系统。当时，规则数写到了将近1万条，都是资深广告领域的业务专家写出来的。后来，我们用机器分析数据，最后写出了1000亿条广告规则，而且比人写的1万条更精细。所以，经过四年，收入提升了8倍。”

火车已然高速运行，就别再执念你的宝马良驹了。

不管你认不认可，机器学习大面积取代人工判断的时代，来了！

训练机器才重要

先讲个真实的事情。

在和某前辈聊天时：

前辈问：“优化点击率，你做了哪些规则？”

我一脸的茫然：“优化点击率是机器该做的事情。”

前辈追问：“那要策略产品干什么？”

是啊，要策略产品干什么？当然不是去跟机器比执行

效率，而是应该训练机器啊。

训练机器通常需要以下几个步骤：

第一，设定目标。

机器擅长的是数据优化，产品经理的职责是给机器设定合理的优化目标，可以概括为：明确产品场景，确定业务的核心指标，并以灵敏指标拟合长期目标。

以内容推荐产品为例，如果你负责的是一个垂直频道，那么这个频道的留存率和内容消费总量肯定是核心指标。对留存率来说，在产品成熟期不敏感，我们可以通过更容易观测的人均消费量来拟合这一目标。对内容消费总量而言，它等于频道渗透率乘以人均消费量，因此，所有的短线优化都可以围绕灵敏的局部指标渗透率和人均消费进行。

目前，由于机器可以支持更高的复杂度，产品迭代已经逐步从单目标向多目标进化，即在不降低点击率的前提下，优化互动量（分享、评论、点赞），在互动量优化中，优先优化分享量。

第二，制订保底方案。

机器的优化不是一蹴而就的，需要一个不断改进的时间周期。在机器算法尚未完善前，产品经理应该给出一个保底方案

作为基线。

以反低俗为例，在模型识别不好的时候，我们可以通过用户举报反馈、数据异常波动复审等方式拦截一部分低俗内容。以此作为基线保证用户的基础体验，留给机器学习和优化的时间。

第三，发现问题。

与其说是发现机器推荐的问题，不如说是发现产品经理在制定目标时存在偏差的问题。

在机器优化的过程中，往往会出现为了极致优化目标A，造成其他部分有损的情况。比如，目标设定的是优化用户的阅读完成度，那么机器执行后最直接的结果就是只推荐短文（一屏展现完毕，用户点击后就100%完成阅读）。这样的结果显然不是我们想要的。为了优化，我们需要进一步修正目标的设定方式。比如，内容至少要多于多少字，对哪些内容可豁免（如快讯），等等。

好的算法结果，既离不开优秀技术研发的实现，也离不开优秀产品经理的投入。

输给机器不冤，能训练出机器才重要。

数据分析驱动产品迭代

如果不能量化，就不要妄论优化。

随着MVP、growth hacking概念的日益普及，有越来越多的业务在产品迭代中应用了growth hacking的思路。然而在实践过程中，我们也发现了一些问题：设立独立的growth部门，在从业务咨询的角度帮助业务线做分析、提改进意见的时候，难免会出现不接地气、对应用场景缺乏足够了解等业务问题，也会出现随着公司膨胀带来的跨部门沟通协作等管理问题。

更好地解决这一问题的方法，莫过于产品经理和数据分析师角色的融合：产品经理能够更深度地从业务逻辑出发找问题，用数据分析的方法驱动产品迭代。

确定核心指标

“不只为做而做”，产品经理所提出的优化和迭代应该建立在服务核心目标上，即保证所有做的功能，是致力于改进核心指标而不是局部业务指标的。比如，不能为了追平竞品功能，就盲目地给自己的产品画蛇添足。追加的功能，即使功能指标不错，也必须考虑其对核心指标的影响。

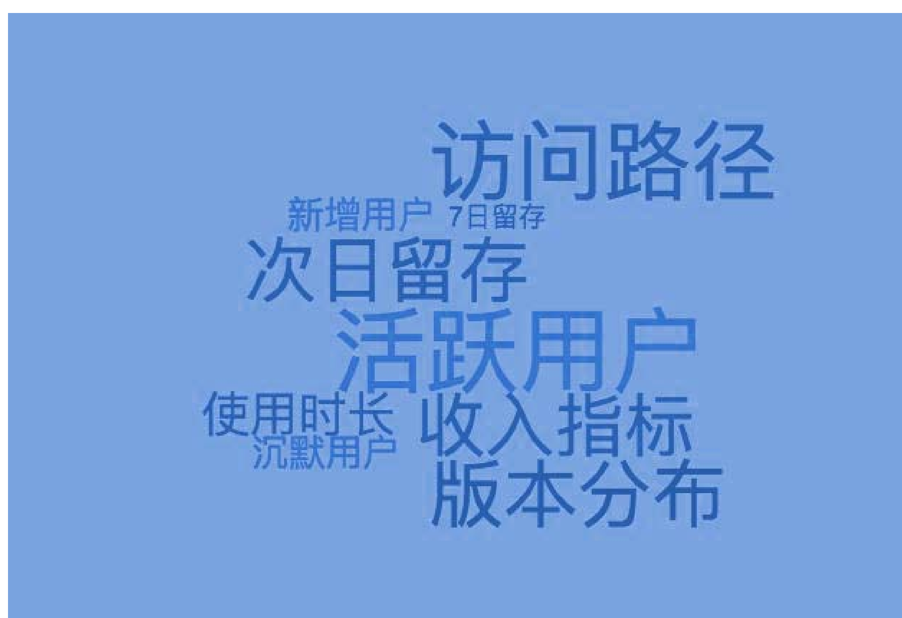


图8–2 友盟后台指标摘录示例

以图8–2为例，图中的各个指标是从友盟后台随意摘录的，在这些指标里，哪个才是一款产品的核心指标呢？

我个人给出的答案是：有价值用户的长期留存。

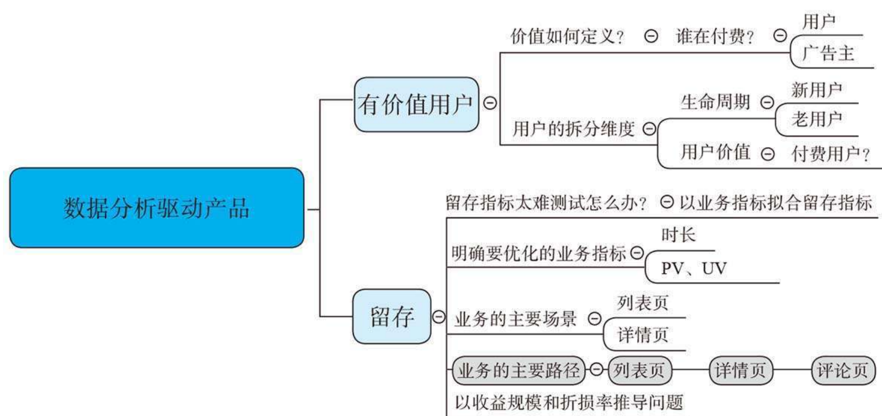


图8–3 产品的有价值用户和留存拆解

下面我们将结合图8–3来做简单的拆解。有价值的用户是指能够最终给公司带来收入的用户，长期留存可以保证收入的稳定性和可持续性。

在价值的定义上，根据具体付费方的不同，可以大致分为C（消费者）端收费、B（企业）端收费，在不同的产品中，其盈利方式和有价值用户的定义会有所不同。

- 以信息流产品为例，有价值的用户可能是那些点击或浏览信息流广告的用户。
- 以游戏产品为例，有价值的用户可能是付费的“氪金”玩家。
- 以直播产品为例，有价值的用户可能是打赏的金主。
- 以职场社交产品为例，有价值的用户可能是活跃在这一产品上的、可被招聘和挖掘的职业用户。

无论收益途径如何，用户价值都会被归结到特定渠道的LTV（life time value，生命周期价值）上，产品层优化应旨在提升用户所带来的全生命周期价值。

核心指标的拟合

尽管我们都认可长期留存是衡量用户产品的核心指标，但是在操作过程中，迭代是快速进行的，无法对每一个方案都开

启几周甚至数月的长期实验以验证这个方案对留存率的影响。

实操中，比较可行的方案是：以敏感的核心业务指标拟合不敏感的长期留存指标，即明确不同产品场景下，和留存指标关系最为紧密的那些指标，后续的产品改进都应该将核心业务指标作为评估标准。不同产品的核心业务指标通常是：

- 对于信息流产品，可能是点击率、停留时长。
- 对于社交类产品，可能是用户的关系数，以及基于这些关系所能产生的消费量。
- 对于直播类产品，可能是观看时长、关注的主播数量。

以领英为例，它是全球最大的职场社交平台，有4亿多名注册用户，一直保持着稳健的增长势头。实验发现：在通过搜索渠道来到领英的用户中，只有接受了好友邀请的用户才会反复访问。此外，用户增加的好友越多，留存的概率越大。通过不断地试验发现，如果一个新用户在第一周添加了5个社交关系的话，这个人就会在领英上沉淀下来，留存率较高。因此，领英将新用户“在一周内增加5个社交好友”作为用户增长的核心指标。

业务的理解

在具体的工作中，每个产品经理分配到的业务是不同的，因此需要对自己负责的业务场景有更深刻的了解，才能分析、产出潜在收益更大的方案。

1. 对主要场景的理解

在自己负责的业务范围内，主要目标是什么？在该业务范围内，用户是怎样消费的，影响用户体验的因素有哪些？

以优化登录率这个任务为例，国内手机软件的登录框提供的登录方式通常有手机号、微信、微博、QQ等。对于这几种方式之间的选择，我们会发现不同手机软件的做法各异：

- 爱奇艺：默认手机号登录，点击“更多登录方式”才提供弹出框选择更多登录。
- 得到：并列提供微信登录和手机登录。
- 今日头条：并列提供手机号、微信、QQ登录。
- 豆瓣：优先提供手机号、邮箱登录，页面下方提供微博和微信登录。
- 腾讯新闻：只提供QQ、微信登录。

如何评价产品对登录方式的取舍，页面元素布局的合理程度呢？这需要按照三部曲：理解现状、提出假想、验证假想的

方式来驱动迭代。同样以登录为例。

理解现状，在于明确当下的业务情况，确定核心指标。

在登录场景下，可以将登录用户占比和单页面登录转化率作为核心指标，作为后续优化的基线。登录用户占比是全局指标，用户登录占比越高，最终结果就越好。单页面登录转化率是一个过程衡量指标，代表引导用户登录所采用方式的有效性。在有些场景下，为了凑最终的业务指标，会采用强打扰、高频露出的方式来进行引导。在这种情况下，尽管最终的登录用户占比比较高，但单页面的登录转化率是较低的。衡量单页面登录转化率，有助于减少强打扰的情况。

确定了核心指标后，还需要围绕业务场景的人群和使用路径进行摸底分析。比如，现有登录方式的使用分布是怎样的；在什么样的场景下弹出登录页面，用户的登录意图更高；不同人群的登录行为有怎样的偏差。对业务既有数据进行详细的分析，有助于我们找到进一步迭代的优化切入点和优化方向。

提出假想，是基于业务的现状，围绕核心指标的优化提出可能的解决方案。比如，如果现在只有手机号登录的方式强露出，其他登录方式是折叠的，那么，将主流的几种登录方式都露出，是否有助于给用户更多选择从而提升登录率？又如，我们发现新用户采用手机号登录的较少，老用户后续关联手机号的情况更多，那么，是否在面向新用户时不用强调手机登录，

改使用门槛更低的联合登录方式？又或者，可以使用赠送业务红包的方式，让用户主动验证手机号？

验证假想，是通过评审后，上线一两个可能性和收益相对更高的方案，来观测局部业务指标的变化（登录率）和全局核心指标的变化（用户的停留时长）。

2.对主要路径的理解

用户是如何到达你所负责的业务场景的，其上下文背景是怎样的？

个人认为，单一页面的优化很容易到达瓶颈，只有结合上下文，才能够带来更显著的提升。登录时，常见的处理方式是，将强依赖于用户信息的产品环节和登录打包，将高级功能和登录打包，将老用户的惯用路径和登录打包。

以滴滴打车为例，主要的使用场景是打车，叫车完成后以“方便司机接您时联系您”为理由，让用户更愿意填写手机信息。类似的，视频、音频应用只有登录才能缓存音视频，在未登录状态下只能收藏一定数量等。



图8–4 滴滴打车验证手机界面截图

3. 根据全局收益进行优先级判断

产品经理所做的每个决策，每个功能迭代都或多或少地涉及研发、QA（质量保证）、运营等环节，既然是消耗公司成本，那么就必须要说清楚收益是什么。

首先，收益一定不能只是局部业务的数字，这一点对于创业公司尤甚。要做对全局有意义的事情，而不是局部有变化的数字。其次，要深挖收益来自哪部分，是改进了低活跃用户，还是进一步促进了高活跃用户。再次，在明确收益的同时，也需要明确是否造成了其余数字的折损。

在跨部门项目上，最容易出现的情况是：A部门第一个月做高了A指标，完成了KPI；B部门第二个月做高了B指标，降低了A指标的增速，也完成了KPI。在两个部门都只关心短期指标的前提下，看似一团和气，其实对全局是不利的。

最后：数据是死的，解读是活的，如何解读，各凭良心。

个性化的好与好的个性化

什么是个性化的好？在既定的指标体系下，借助技术手段不断追求更好的数据表现。

什么是好的个性化？在不改变用户目的的前提下，借助技术手段达成用户效率和体验的提升。

前者是技术流，后者是价值观；前者强调客观标准，后者强调主观目标。这两者在项目初期往往会高度一致，但当项目发展到中后期时，却往往会出现错位。

以搜索为例。

好的搜索应该做到用完即走，以优化用户的点击满意度为己任。从理想状态来说，搜索引擎应该只提供一屏幕结果，在第一屏就能够满足用户的个性化需求，让用户用完即走。

但在实操中，为了追求搜索量或停留时长的指标，搜索引擎越来越多地插入了“相关推荐”“大家还在搜”的引导，以侵入式的体验换取搜索量和停留时长的增长。

对于内容消费，同理。

个性化的好，是不断追求数据指标。在技术水平不足以支撑目标的时候，就越来越容易围绕马斯洛需求模型的塔座转圈。所以，我们才看到网络上会有那么多调侃“××惊讶体”“××爱国体”的内容。

而好的个性化，是需要克制欲望的，需要引入更多维度的人工评估进行决策权衡，以可量化的短期指标损失来保证不能轻易量化的消费体验。

个性化的好，是生意。

好的个性化，是理想。

本书由“ePUBw.COM”整理，ePUBw.COM 提供最新最全的优质电子书下载！！！

Part 02
关于自媒体

本书由"ePUBw.COM"整理，ePUBw.COM 提供最新最全的优质电子书下载！！！

好内容为什么没人看

在与运营人员沟通的过程中，我最常被问到的问题是："这个内容挺好的，但为什么推荐量和阅读量都很低呢？"

我充分相信运营和BD（商务拓展）人员的判断：对于内容的质量，文化人比我这个工科男更有话语权。那么，就让我们尝试分析一下，好内容为什么没人看。

第一，质量不等于销量。

作为阅读者，看到好的内容推荐不出去，固然有遗珠之痛。

但作为运营者，还是要跳脱读者的角度，站在相对客观的角度去评价一个内容的消费普适性。如果这篇内容本身就是深度阅读内容或题材偏小众（比如爵士乐），那它的消费量或许理当如此。我们可以将其与文艺片类比，摘录一段报道：

> 明星云集的文艺大片《黄金时代》最终总票房仅5151万元；《推拿》获得第64届柏林电影节银熊奖等奖项，但国内票房只有约1290万元；《闯入者》《念念》《心迷宫》等文艺片总票房都是1000多万元。

票房低至不可思议的，是具有文艺元素的传记片《启功》，总票房只有约40万元，差不多等于目前市场上启功书画作品一平方尺的价钱。口碑好、票房惨、骂院线、怨社会，不了了之，这似乎成了文艺片的宿命。

如果一篇内容是内向性创作的，是写给自己的，那么这篇内容我们更应该从艺术品的角度，而不是从消费品的角度看待。其追求的首先应该是给少数知音群体更深的触动，而不是作为大众消费品评价标准的10万次以上的阅读量。同时，在评估这类内容的阅读量时，应该与其同类型内容、同作者内容对比，而非泛泛地评价其推荐量的高低。

只有一篇内容从题材选择上是具有消费普适性的，其阅读体验才可能是相对较优的。对于这类内容，我们再来深究其没有被推荐好的原因，分析是内容包装的问题（标题、封面）还是推荐问题（内容识别、冷启动、用户阅读指标）等。

第二，与其抱怨没有推荐量，不如来谈谈怎么推荐得更好。

推荐系统是一个会极度放大优缺点的系统：点击率高、阅读体验好的内容，其传播度会被极大地放大；点击率低或者点击率高但是阅读体验差的内容，其传播范围会非常快地收敛。

好内容的作者通常已经是被系统标注的优质作者，在冷启

动过程中，这些作者由于有更好的历史表现，已经得到了远超普通作者的冷启动曝光量。当冷启动过程结束时，就变成了内容级别的点击率和阅读体验的公平较量。

我们可以将内容推荐想象为一场体育比赛，黑马可以爆冷夺冠，名将或许可以免于预选赛，但还是有可能失去水准止步半决赛。所以，并不是优质作者的内容就一定会有高的阅读量。

通常，我们不建议用超过10万次的阅读量这样的标准去简单粗暴地衡量内容。我们可以将内容的阅读体验指标化、可衡量化，以进行进一步的分析，可以切入的角度如：

- 点击率：酒香也怕巷子深。如果点击率太差，先别废话，回炉改封面和标题去。
- 阅读反馈：点赞、评论数与阅读量的比例，是否与其过往的内容一样？还是降到了平均线以下？如果是，那说明要么是这篇内容质量有问题，要么是推荐系统推荐的目标人群不对。

当越来越多的内容分发平台开始引入机器推荐的时候，自媒体也许需要了解不同平台的分发规则，让有价值的内容获得应得的流量，进而让自己的品牌有更高的价值。

让好的内容有人看，让好的创作者有钱赚——这就是本书第二部分的主题。

本书由“ePUBw.COM”整理，ePUBw.COM 提供最新最全的优质电子书下载！！！

自媒体的数据分析

“If you can’t measure it, you can’t improve it.”

“如果你不能量化一件事物，就无法改进它。”这是一句在增长黑客圈非常有名的格言，套用自管理学大师德鲁克的名言“You can’t manage what you can’t measure”（你不能管理无法量化的事物）。我们在开展业务的过程中，需要首先对业务进行量化（指标化）处理，只有可量化的业务才可以被更好地分析和改进，不会深陷“我觉的做得好，老板觉得还不够，大家觉得还不错”这样表述模糊不清的泥沼当中。

这一章会聚焦自媒体应该如何分析内容传播和粉丝增长过程。我希望能够有一种途径可以量化一篇内容在不同平台的分发和受众反馈情况，用数据更好地辅助创作者稳定持续地产出80分以上的作品，在各个平台获得稳定的流量。

鉴于我在自媒体和今日头条的工作经验，我会分别以微信公众号和头条号为例进行说明。

他山之石：BuzzFeed简介

“他山之石，可以攻玉。”数据分析辅助内容创作在国外最典型的样本非BuzzFeed（一家美国新闻聚会网站）莫属了。

员工总数逾千人，拥有全球新闻团队、自己的视频制作工作室、尖端数据运算中心和内部创意广告机构，每月全网超过50亿次阅读——这就是BuzzFeed，怎么看都是一个庞大的媒体集团。

然而，当我们深入了解BuzzFeed的工作流程后就会发现，与其说它是一家媒体公司，不如说它是一家科技公司。“数据驱动内容创作”，是BuzzFeed给自己贴上的标签。

在这家公司的内容创作和分发过程中，反馈闭环（Feedback Loop）是出现频次最高的术语：将市场环境和读者反馈数据尽可能地量化和结构化，以及时反馈给运营人员、内容编辑，从而构成一个辅助创作的闭环。

数据收集与处理

由于很早就树立了“将内容分发到用户常驻的平台，而非吸引用户到自己的平台看内容”的思路，YouTube、脸谱网，推特、Vine（一款视频分享移动应用）等一众平台都成了BuzzFeed的战场。为了更好地理解用户反馈，BuzzFeed追踪收集并汇总了各个平台上的阅读传播和互动数据。比起阅读量、分享量这些结果指标，其更关心内容分发和传播的过程。

以图10–1为例：如果只看终态数据，那么报表展示给你的无非是一篇内容从不同的平台渠道上获取了多少流量而已（如图10–1左）。但是深究下来，这些流量是如何而来的呢？是通过脸谱网或推特上的分享，还是通过电子邮件的传播？哪些节点带来了更大的辐射量和扩散量？只有深入分析传播路径，才能给出这些问题的答案（如图10–1右）。

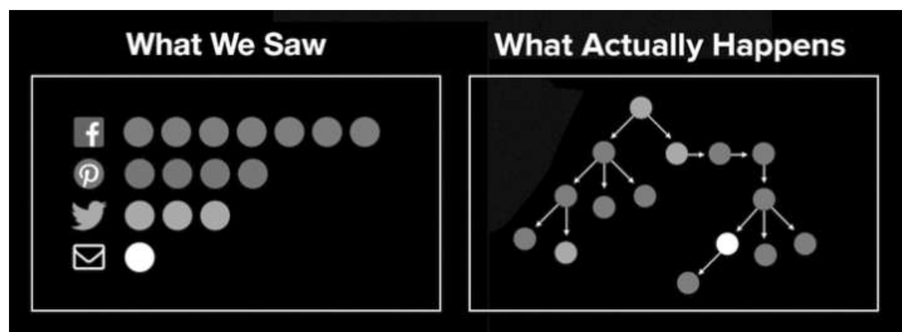


图**10–1** 数据收集与处理示意图

不止于结果，BuzzFeed开始探究一篇内容的传播轨迹，开发了一套名为POUND（Process for Optimizing and Understanding Network Diffusion，网络传播优化及理解流程）的系统，通过给同一页面的不同分享追加不同参数的方式，来追溯内容在网络上的传播过程。

通过这套系统，人们意识到：传播并不是一棵简单的传播树，而是一片密集的传播森林。每一次分享都构成了一棵新树，如果一个人的社交好友很多，那么这个节点的一度关系触达就会很广。如果内容的质量一般，没有引发进一步的传播，

那么就会止步为一棵低矮的树；如果内容引发了大量传播，就会变为一棵繁茂的大树。

因为社交传播对内容的影响力贡献巨大，BuzzFeed在衡量内容价值时，会更看重社交分享所带来的阅读量，并以病毒传播提升系数（Viral Lift）指标来衡量这一过程。

对粉丝已经超过百万、内容阅读量篇篇超过10万次的自媒体大号来说，这一衡量指标的制定不难理解：如果一篇内容的阅读构成主要来自自己的粉丝，那么就意味着粉丝们更有可能是基于惯性打开的，但是由于内容并没有触及他们的兴奋点、没能引发认同，从而失去了进一步扩散传播的可能，也失去了触达更多潜在粉丝的机会。

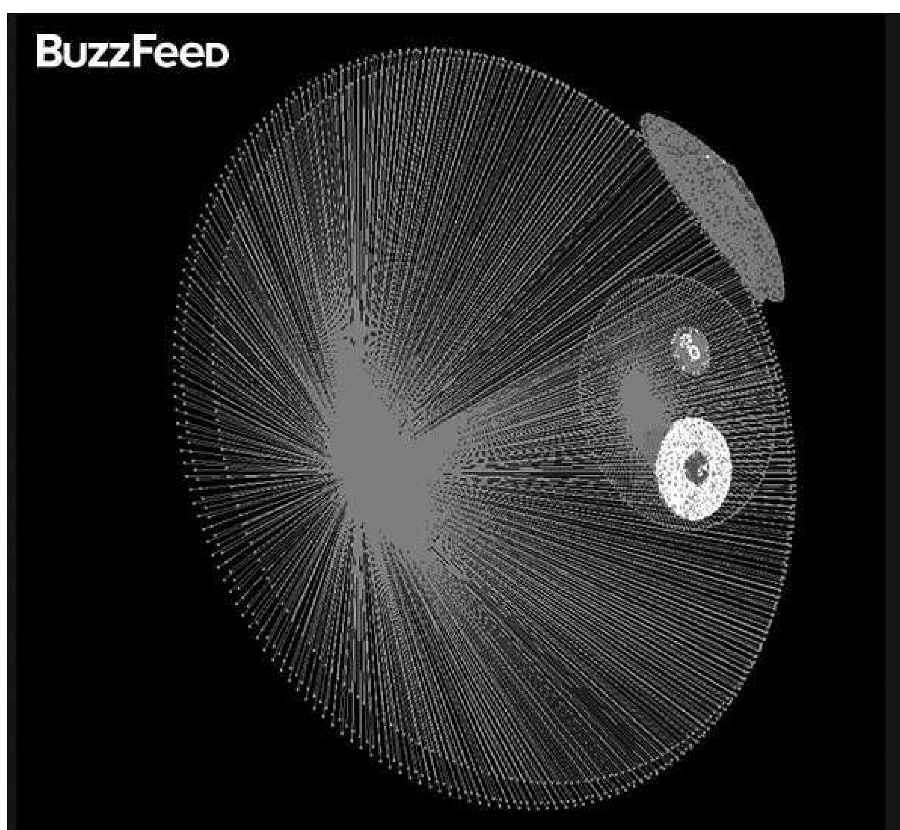


图10–2 BuzzFeed POUND系统截图

对BuzzFeed而言，在总阅读量相近的情况下，一篇病毒传播提升系数更高的内容，因为收获了更多的社交传播而变得更有价值。有了技术追溯的支持，内容团队才能够总结出不同平台上的内容偏好和病毒传播模式。

人们往往会夸大BuzzFeed的“内容三板斧”——清单体、短视频和小测验，却忽视了在这些内容背后更深层次的数据支持。

降低数据使用门槛

有扎实的数据基础只是第一步，如果一线的编辑、商业化人员不能运用这些数据，数据收集工作仍然是徒劳的。为了降低数据的使用门槛，BuzzFeed自研了20多款内部数据分析工具，以满足不同部门的数据整理和查看需求。

对于不熟悉编程语言的员工，能够在其Looker数据平台上以可视化的方式查看不同来源的数据，直观地了解内容的传播轨迹和传播效果。无论是文字编辑、视频编辑，还是业务分析师、商业广告支持人员，BuzzFeed希望员工都能够更清晰地认知到自己的每一项工作所产生的数据反馈，及时了解目标受众到底关心什么、究竟喜欢什么，按需调整自己的工作策略。对数据的重视和应用像基因一样，已经嵌入BuzzFeed的每一个部门。

除了基础的后验性阅读追溯数据外，BuzzFeed还致力于通过数据优化创作过程本身。比如智能封面图截取工具，能够根据不同平台的特点、画面的清晰度与表意性自动截取（在视频上传后的系统封面候选集合中，也运用了类似的思路）。又如实时热点监控工具，通过监控社交网络的潮流变化，可以及时将当下流行的爆点推送给自己的内容编辑等。

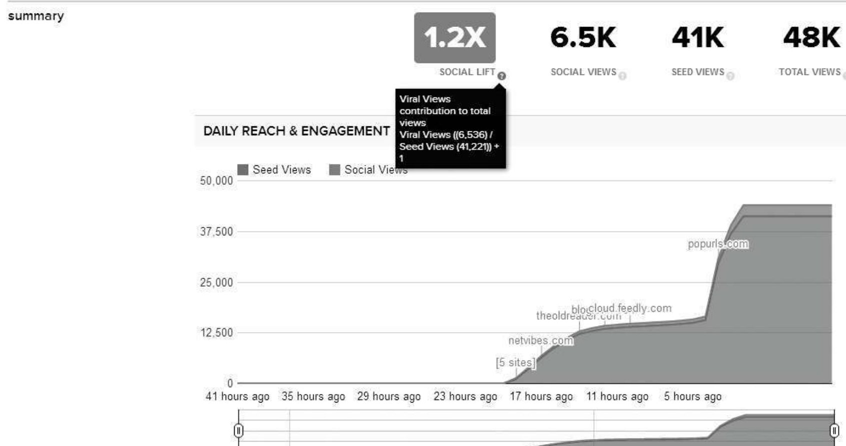


图10–3 BuzzFeed数据平台

2016年2月，《快公司》杂志公布了全球最佳创新公司50（Most Innovative Companies 50）榜单，BuzzFeed位列榜首。BuzzFeed以数据驱动媒体的方式或许可以给我们借鉴：内容创作和内容分发可以是数据化的，科技的辅助应用或许能给变革者带来超越其他媒体前辈的机会。

内容阅读分析

为了获得持续稳定的内容传播量，我们需要结合平台所提供的数据进行分析，研究其在不同分发平台上的传播特点。在这一节中，我们将其分为一次打开和阅读体验两个部分进行探

讨。

一次打开

一次打开是指不依赖社交传播和搜索传播所带来的阅读量。对于微博微信等粉丝分发平台，是指粉丝带来的阅读量；对于今日头条、快报等推荐分发平台，是指经由推荐带来的应用内阅读量。

影响内容一次打开率最直接的因素有两个：活跃粉丝量和内容包装的好坏。

无论是粉丝分发平台还是推荐分发平台，活跃粉丝量都会影响内容展示的基本盘，只是传导系数不太一样罢了：对于粉丝分发平台，有多少活跃粉丝就意味着有相应比例的基础展示量；对于非粉丝分发平台，自媒体账号的活跃粉丝量会被视作其平台的贡献程度和受众情况从而影响分发权重。同样的一篇内容，10万活跃粉丝的账号发布一定比1万活跃粉丝的账号发布获得的基础展示量更大。

内容包装影响的是从展示量到阅读量的转化过程，更好的标题和封面图就像精致的产品包装，有助于更好地促成转化。

对于标题，我习惯引用咪蒙的观点：“不能在一秒钟看明白的标题，不适合传播。不是读者理解不了，而是他只能给咱们一秒钟。”

在咪蒙的工作方式中，她会把当天写好的内容发出来，交于自媒体小组的员工开始取标题。15~20分钟内，每人至少取5个标题。这样，每一篇内容背后，都有近100个标题！咪蒙随后会从中挑出5~6个，放到3个用户顾问群里投票（每个顾问群都有人负责统计票数）。咪蒙会参考最终的投票结果，决定用哪一个标题。这种方式，其实就构建了一个最朴素的AB系统并以此验证标题对内容一次打开率的影响。

头条号的双标题、双封面，提供的就是优化点击转化的可能性。使用这一功能，相当于自媒体拥有两次冷启动机会。图10–4就是头条号后台对双标题、双封面内容的展示方式，从中我们可以看到同一篇内容在两个不同的标题、封面组合下，阅读量存在着显著差异。



图**10–4** 头条号后台对双标题、双封面内容的展示方式

为了便于比较，我们通常计算一次打开率而非打开量。对于粉丝分发平台，一次打开率的计算公式为“一次打开率＝阅读量/粉丝量”；对于推荐分发平台，一次打开率的计算公式

为"一次打开率＝应用内阅读量/推荐量"。一次打开率对于不同分发模式的平台有不同的意义：

- 在粉丝分发平台上，与阅读量相比，一次打开率能更好地衡量自媒体账号的价值。为了保证更高的一次打开率，需要更科学地获取精准粉丝、重视粉丝群体的互动维护，使自身能够始终保持较高的粉丝触达率。
- 对于推荐分发平台，一次打开率考量了自媒体对平台分发特点的认知和对内容包装的功力。一次打开率（推荐点击率）高意味着具有成为爆款的可能，甚至会收获比粉丝分发平台更大的阅读量。

阅读体验

如果说一次打开代表了上门推销人员成功敲开了用户的房门，那么，阅读体验衡量的就是用户是否会为这次推销最终买单。

用户是带着对作者品牌和对标题封面的认知预期打开内容的，只有实际消费体验与用户的预判一致甚至超出预判，才算得上是好的阅读体验。我们之所以抵触标题党，并非因为它骇人听闻，而是因为用户抱着骇人听闻的预期，点击进入详情页之后才发现内容平淡无奇。

想要衡量消费体验，内容的消费比例是一个重要指标。对

图文来说，这个指标是平均阅读进度和阅读速度；对视频来说，是播放进度和播放时长（因为用户会有快进的行为，所以100%的播放进度并不意味着100%的播放时长）。平均消费比例越高，代表用户的认可度越高、消费体验越好。参考这一指标通常会给我们很多有意思的发现：

- 片头：很多短视频综艺节目都会有自己的片头，越长的片头通常会带来越高的跳出率。类似前贴片广告，在用户对内容品牌没有强认知的情况下，片头实际上降低了他们获取内容的效率。一般建议将节目品牌插入内容播放过程中，作为转场画面以获得品牌曝光量与播放量的平衡。

- 片尾：在内容质量尚可的前提下，片尾广告的触达率通常都不错，如何提升转化率，考量的就是自媒体的引导性内容制作功力了。由于看完了整个视频，所以自媒体在片尾引导用户关注自己的效果要比片头好一些。

- 相同字数的内容、多图内容的平均阅读速度一般略快，这可能是因为用户的注意力主要被图片吸引，从而更容易略过文字的部分。

- 对于推荐算法分发的系统，跳出率会存在一定的偏差。比如，标题中使用了歧义词导致系统将其分发给了非目标用户，这部分用户的头部跳出率通常很高。

图10–5为图文和视频的消费完成情况分析。在一些垂直网站，如优酷、喜马拉雅，对内容的跳出点给出了更为详尽的分析数据。

平均阅读进度	跳出率	平均阅读速度
86.6%	15.27%	16字/秒
平均播放进度	跳出率	平均播放时长
57.51%	9.54%	00:01:32

图10–5 图文和视频消费情况分析

提升内容消费完成度考量的是自媒体内容创作的真功夫，只有引人入胜的内容，才能实现更有效率的信息传达，收获更多的数据。

除了内容消费比例指标外，内容是否引发了读者互动也是一个常见的考量角度，如评论、点赞、收藏、转发等，这些指标的横向对比通常用于组织内部考核不同编辑的创作能力。其中，我个人建议需要额外关注的是评论和转发两个指标。

评论扩展了内容的深度，更多的用户评论给正文提供了不同的视角分析和信息补充，可以有效地引发围观用户的阅读兴趣，提升用户在内容页的整体停留时长。促进评论的提升是有技巧的：选题有冲突性或者有代入感，在正文中留有悬念，主动引起讨论、投票，在评论区中带节奏等都是行之有效的方法。

转发扩展了内容的广度，更多的转发能够带来更多面向潜在受众的曝光，这一点对于以粉丝分发为主要模式的系统尤为重要。对于转发指标，我们可以套用BuzzFeed病毒传播系数的计算方式：Viral Lift ＝ 1 ＋ 传播阅读量／一次阅读量。

粉丝增长分析

对自媒体而言，虚高的粉丝数除了忽悠投资人和广告主以外，对自身成长是毫无意义的。想要可持续发展，只有精准的粉丝才有价值。那什么是精准的粉丝呢？我想，应该是能够给自媒体带来变现价值的粉丝。

粉丝增长分析，就是围绕有变现价值的粉丝算的一笔账：如果一个粉丝能给自媒体带来5元的净利润，那么花低于5元的成本来购买这个粉丝就是划算的。

首先，我们引入三个概念：LTV、CAC和ROI。

LTV：一个粉丝从关注该自媒体的那一天开始到取消关注或者脱离平台，能够带来的收益，通常按渠道来计算。如果某个渠道或某个平台的用户不精准、付费意愿差，那么这个渠道的用户LTV就相对较低。

CAC（Customer Acquisition Cost，用户获取成本）：同

样跟渠道有关，自媒体需要持续发现低价、优质的渠道，抢占红利期。比如，2015年年初，微信广点通渠道的公众号涨粉，单个用户获取成本只有1～2元。

ROI（Return On Investment，投资回报率）：其计算公式为ROI =（收入–成本）/成本。应用于粉丝增长场景就是ROI =（LTV – CAC）/CAC，衡量的是自媒体从一个粉丝身上挣到的钱是否能够覆盖获取这个粉丝的成本。

通常，我们应该做“ROI＞1”的事情，这样才能够保证业务的可持续发展。后续讨论也都从这个角度展开。

首先，我列举一个亲身经历的事情。2015年年初，微信朋友圈集赞的路数还很火热，微信公众号“凯叔讲故事”也不能免俗，搞了三场集赞有礼的活动。由于当时凯叔讲故事是拥有商品消息功能的，即发布的图文点击后可以直接跳转至某个站外链接，我们可以有效地监控不同用户的访问情况。

三场集赞有礼的活动，表面上看起来很光鲜，各种指标在活动日都有显著的提升。但是兴奋过后的分析，却让人笑不出来。以下是当时所进行的数据分析（数据有调整，但比例和趋势不变）：

表**10–1** “凯叔讲故事”三场集赞活动数据分析

A 活动名称	B 活动时间	C 参与人数	D 与前一次活动重复	E 发起人重复比例	F 拉新人数	G 次日残留人数	H 次日残留率	I 次日活跃人数	J 活跃比例	K 人均拉新	L 任务量
活动 1	9 月 25 日	8800			38222	34587	90.49%	1280	3.70%	4.34	15
活动 2	10 月 20 日	8614	1294	15.02%	38105	32429	85.10%	596	1.84%	4.42	20
活动 3	11 月 11 日	4071	798	19.60%	14550	12189	83.77%	212	1.74%	3.57	25

首先简单介绍一下三次活动的规则。只有粉丝才能够发起活动，通过社交传播集赞来获得实物礼品（电影票、玩具等），只是赞的门槛略有不同：

- 活动1，收集15个赞。
- 活动2，收集20个赞。
- 活动3，收集25个赞，且一个用户只能够投3票。

结合表10–1中的数据进行分析，从拉新角度看，在第三次活动的时候发起人的重复率已经近20%，说明有1/5的人都是同一批用户。随着集赞活动的不断推进，同一批老用户能够带来的新用户量必然越来越低。正因为如此，在第二次活动将集赞指标加码33%的前提下，人均拉新量只有略微提升。第三次活动将集赞的指标加码了25%　且人均只能投3票的情况下，人

均拉新量还是降低了近20%。

从用户质量角度看，我们考虑次日残留（参与活动后没有取消关注的人数）与次日活跃两个指标。从表10–1不难看出，两个指标都在不断下降。如果以活跃用户数来核算成本的话，已近几十元了。要知道，彼时微信广点通买用户只有几元成本。

于是，我们开始反思用户为什么要参加集赞有礼的互动？唯利罢了。

老用户发起活动，他所带来的新用户无非是看一眼品牌广告，帮朋友支持一下。如果实物奖品价值还不错，那么这个新用户也会发起，但目的在于奖品而非公众号。此外，设计过高的门槛使得有的用户不乐意参与，有的用户则自建了群开始互相点赞，达不到拉新的作用。

想要提升用户增长的ROI，就必须降低新用户中非精准用户的比例，从而拉升渠道用户的LTV，降低用户获取成本CAC。

由此入手，我们优化了转发集赞拉新的方式，新一期集赞活动的礼品是凯叔西游记第二部的部分章节收听权（如图10–6），每个用户只需要拉5个用户就能获得。因为是虚拟产品，CAC得到了显著降低。而儿童故事音频跟微信账号捆绑只能自

用，用户如果不是真有需求就绝不会参与活动，从而提升了新增用户的精准程度。在后续的此类活动中，我们都有意识地对礼品形态进行控制：实物礼品用于鼓励已有的活跃老粉丝，虚拟礼品用于激活和拉新。









图10–6 凯叔讲故事“西游记”章节收听页面

运营微博、微信这样的平台，我们很容易沉迷于追求短期粉丝数字的飙涨而举办各种活动。某些活动拉新从CAC的角度看或许是划算的，但深究其带来的真实活跃粉丝，就会发现其ROI往往远小于1，充其量只能算是一次展示广告曝光，做了一笔又一笔的蚀本生意。

想要算清楚用户的ROI，有赖于用户身份的识别和渠道的追溯。在这一点上，微博和微信的基础建设比较完善。以微信为例，基于平台提供的二维码，我们可以追溯不同用户的来源并进行标识，就像是安卓应用下载对不同的应用市场打不同的渠道包一样。进而，通过平台提供的外链功能、私信功能，来识别用户的活跃情况。

除了微博、微信以外的分发平台，用户身份大多不可追溯，这种情况更类似于AppStore（苹果应用程序商店）的推广。你需要知道哪个渠道是真正有效的，这样才能够改进后续的步调。比如，可以针对不同渠道设置不同的回复关键词，通过参与比例来预估不同渠道的转化情况。

如果采用平台内自媒体互推换量的方式，可以首先估算出自己内容的转粉率，再与其他自媒体换量之后进行计算：通过其他自媒体互推带来的粉丝量＝当日总增长粉丝量–自有内容的增粉丝量＝当日总增长粉丝量–自己内容的曝光量×转粉率；或者，邀约其他自媒体推荐自己的一篇无时效性内容（如

工具型内容，汇总了一些高热文章、提供了资源下载等），那么这篇工具型内容的阅读量增长也能作为辅助计算因素。

对于付费增粉渠道，每隔一段时间暂停一下，比较渠道暂停前后自己粉丝量的增长情况和粉丝阅读占比、线上活动粉丝参与量情况，以便对付费渠道有更清楚的了解。

本书由“ePUBw.COM”整理，ePUBw.COM 提供最新最全的优质电子书下载！！！

内容快销：标题党的二三事

曾经，某平台在追求点击率的路上用力过猛，导致推荐了不少标题党的内容，也因此收获了“某某震惊部”这样一个戏谑的名头。更有人笑称：“某某震惊体”“某某脸红体”“某某中国体”堪称自媒体运营的入门必修课。

比如，有这样一个段子：

> 一场中美选手参加的比赛，中国选手战胜了美国选手，这事儿放到某平台，标题应该是什么？
>
> 参考答案：中国人打美国，暴爽！没Wi–Fi也要看！不看不是中国人！

这一节，我们就聊聊标题党。

标题党的开山鼻祖

我所认识的大部分由纸媒或门户网站转型做自媒体的朋友，在创业伊始往往是排斥标题党的。对此，我通常直言不

讳：你们占据了太久流量分发的位置，从甲方换到乙方都不会干活了。

纸媒标题党的鼻祖，或许应该追溯到1985年。那一年，一本名叫《知音》的纸媒在武汉诞生，由此开创了一个名叫“知音体”的时代。还记得那年的报亭，一本以时尚女性作为封面的杂志在众杂志中脱颖而出，以硕大的字体呼喊出其本期的内容：

- 《再大的恨放下吧，唤醒前夫赢得亲情一片天》（2007年第35期）
- 《风之谷啊我的妹妹，哥哥的未来献给你》（2007年第7期）
- 《外婆杀孙啊可怜的疯狂，血案几多亲情泪》（2008年第18期）

华丽的辞藻、哀怨的情绪、充满戏剧性的冲突，成了知音体标题最鲜明的特点。

在当时的时代背景下，在大量依靠政府拨款订阅的期刊中，依靠零售市场自负盈亏的《知音》杂志宛若一股泥石流，来势汹汹，迅速吞没了中国的二三线城市。可以用一份数据来证明《知音》的辉煌：《知音》是排行全球前五强的著名大

刊，也常年稳居中国杂志发行量第一的宝座，月发行量一度达到700万份。

对于知音体标题，《知音》杂志的副主编夏志军曾在2014年做出过公开的回应：百姓需要这种直白的表达。

自媒体：在商言商

时光轮转，进入了自媒体时代。内容的分发可以说更加市场化了，大量的信息涌现在用户面前，比琳琅满目的书报亭有过之而无不及。对自媒体来说，无论是做GQ（《智族》杂志）还是《故事会》或《知音》，都需要在用户给你的那一次回眸中留下自己的烙印。

媒体是门生意，一门以内容为主要流量货品的生意。其基本的角色就至少应该划分为：生产者和销售者。

内容创作者（生产者），负责制造内容、保证内容质量，他们需要做到：用户读到第一行文字之后就被吸引，带着很好的内容体验完成阅读。考核内容创作者的，应该是内容的读完率、互动率（评论、点赞、转发等）之类的指标。

在内容创作者之外，自媒体还应该有一个宣发者（销售者）。他们负责内容的标题、封面，内容的开篇、结尾，同阅读者互动等。此外，由于不同的平台渠道对内容的分发方式是有差异的，故而宣发者应该熟稔各个平台的分发规则和反馈渠

道：在算法分发的平台上要强化标题的吸引力，在人工分发的平台上还要强调公关能力。

对于熟识的朋友，我一再建议：即使团队再小，也应该设置一个负责内容宣发的职位。因为只有这样，它才能够站在市场的角度重新审视内容，不夹杂创作情感的包装和营销，从而实现内容商品属性的最大化。

自媒体界的“标题党”典型莫过于咪蒙了，以下是她的创作谈摘录：

> 创作最快也得3小时，因为写完通常都要改一到两遍。我一般起标题都至少要半小时到一小时。最长写一两天，改好几遍，还写成一坨屎。阅读量最高的内容是《致贱人》，阅读超过520万次，这篇属于情绪到了写得蛮快的，写一稿花了两小时，后来起标题、修改花了1小时。其实我不算快手，写得慢的比如《如何谋杀你的妻子》这篇，怕细节出错，反复确认，反复修改。不能在一秒钟看明白的标题，不适合传播，不是读者理解不了，而是他只能给咱们一秒钟。

或许你会质疑咪蒙文字的价值，会抨击她内容刻意的导向性。但是，我想提供另一个维度的看法：咪蒙做了十多年记者和编辑，文学底蕴不见得比你差，但为什么选择用这种方式来

运营她的公众号？或许是因为，她明白自己所做的是一门内容快销品的生意！

说出“我的本色是悲凉”的马东，正带着他的米未传媒热情地拥抱着大众。那个在《奇葩说》里不厌其烦地进行花式广告口播的男人，你可还能看得到他当年做《有话好说》《文化访谈录》的影子吗？

既然在商言商，就请别藏着掖着，从起一个挑逗用户情绪的标题开始吧！

内容平台：遏制的其实不是标题党

说完了自媒体角度，再来聊聊平台角度。了解平台的立场，有助于自媒体更好地把握日常内容运营的步调。

我一再强调，平台之所以重视标题党原因在于保障阅读体验。如果用好的标题把用户吸引进来，却没有给其应有的体验，长此以往，平台留存难以维系。从保证用户阅读体验的目的出发，能够避免平台采用非黑即白的态度对待标题党，就有了更弹性、灵活的操作方式。

第一，用滥了的标题党，规则直接封禁。

“惊呆了”“不得不看”这种词汇可以说是用滥了的标题党。由于它们在全网的过度传播，在读者心中，已经基本与“低

质”画上了等号。就算内容真的是不得不看，用户也不相信了。

对于这类标题党的遏制，可以类比对过度广告的遏制。在新《广告法》中，“最”字头、“国”字头以及“唯一”“首个”“领先”等极限用语也被一刀切地禁止使用。对于此类标题党情况，平台可以采用一刀切的管理方式，直接封禁。

这种一刀切的词汇禁用，是不是有误伤？

当然有。

但是考虑到误伤范围和可杜绝的负例，一刀切成了相对可接受的保底方式。只要封禁词库相对可控，就不至于过度伤害内容创作的自由度。

第二，歧义、蹭名人的标题党，借助阅读反馈发现。

封禁了“最”之后，广告里还是会有似是而非的用语；封禁了用滥了的标题党，投机的作者还是会以歧义、蹭名人的方式钻空子。以如下两个丧心病狂的标题为例：

- 美食标题：男生下面好吃，还是女生下面好吃？（讲述的是煮面条）
- 科技标题：让马云和马化腾都动容的人，这个项目遭BAT哄抢（正文与马云、马化腾完全无关）

这样的内容可以通过阅读分析、阅读反馈的数据尽早发现和杜绝。

蹭名人的内容，可以借助文本分析来初步了解正文和标题的实体匹配度是否足够；标题歧义的内容，由于用户的阅读预期和内容差异较大，通常阅读完成度并不好；一些涉及性暗示的标题，可以通过收集和统计样本来训练模型。

第三，基于作者的过往历史，提供不同授信空间。

再次强调，平台遏制标题党目的是为了“保证阅读体验”，只要阅读体验足够好，那么在标题、封面层上的炫技是可以接受的。

因此，在同一个平台上，历史表现不同的作者会拥有不同的授信空间。

在系统识别出标题党的概率后，对不同作者的处理方式会有尺度上的松紧之分。劣迹作者可能干脆就不会获得推荐量，优质作者则可能不会有太多降权，甚至会有人工复查的机会。

把好的阅读体验带给更多人，站在这一点上，平台和自媒体的利益诉求是一致的。

推荐平台优化：从SEO到REO

如果你经历过PC互联网时代，就一定不会对SEO（Search Engine Optimization，搜索引擎优化）感到陌生。

在PC互联网时代，搜索引擎是每个网站不可或缺的流量源头。在特定的搜索词下，更好的搜索结果展示位就意味着更高的搜索流量。为了从搜索引擎获得更多的流量，网站运营者总结了各种各样的优化方式，根据其是否利用了搜索引擎漏洞而被分为“黑帽SEO”和“白帽SEO”。搜索引擎公司也秉持着堵不如疏的心态，尽可能地开放自己系统内可被人理解的信息抽取规则、内容排序规则，以提升业务的透明度，与网站保持信息通畅、良好互动的关系。

理想态下的白帽SEO，是在网页内容有价值的前提下，提升网页内容对搜索引擎的友好度，并逐步通过自有品牌的建立，获得越来越多不依赖于搜索引擎的品牌认知流量。

对推荐引擎而言，其理亦然。内容就像是网站的网页，可以优化内容的构成与包装方式，让它适应不同平台的推荐分发规则从而得到更高的曝光量，我们姑且称之为REO（Recommend Engine Optimization）吧。

让推荐引擎正确理解内容

看过了之前的内容，相信你已经大概了解了一篇内容在经历了哪些过程后，才能被推荐引擎展示给用户。自媒体应该认

识到的是，自己的内容首先需要过机器这一关，然后才能够被分发给用户。只有让推荐引擎正确地理解内容，才能更有助于内容的精准分发。

对图文内容来说，其实相对简单，在大部分情况下只要正常写作即可。推荐引擎会根据正文内容进行文本分析从而判断其主要的内容分类和话题，以进行后续推荐。在涉及一些英文内容的时候，要尽量补充对应的中文译名，以提升机器的理解程度。

对于图文内容，坊间流传可以通过增加热门关键字的方式或是将热门关键字以图注的形式追加进文章的方式误导推荐引擎，以便让推荐引擎将其推荐给更广泛的人群。这种手段比较像早年在网页中堆砌关键字或者背景色关键字的方式，对一些处于初期的推荐系统确实有一定作用，可一旦相应的识别规则上线，将直接影响对应自媒体账号的可信度。

对视频和图集来说，仅拍好内容是不够的，还需要更完善的信息补充。尽管现有的技术可以做一些基于帧的图片分析或视频音轨分析等，但尚达不到大规模工业应用的程度。在这种情况下，自媒体应当尽量完善平台提供的可补充文本信息的字段，如标题、摘要、话题（关键词）等，给机器提供一个更好地理解内容的途径。也正因为视频上附加的文本内容比较少，所以我个人不建议在视频上过分追求双关或者标题的精妙，这

容易使推荐引擎错误地理解你的内容，导致冷启动失败。

让推荐引擎信赖你的品牌

树立品牌的可信赖度，是一个长期且回报稳定的事情。由于每一篇内容都会经历冷启动过程，可信的自媒体品牌会得到更高的授信曝光额度。类似在搜索引擎中，对于两篇正文内容一模一样的网页，会优先展示高权重网站的内容，推荐引擎也会在内容话题相似时优先考虑品牌更好的自媒体。

在前述章节中，我们提到过平台对自媒体的一些考核指标。自媒体如果想要优化分发，应该着眼于推荐系统的长线利益诉求，而不是沉迷于当下的短线推荐规则。因为所有规则都会不断迭代，并最终趋近于推荐引擎的目标。只有知道了推荐系统想要什么，才能够做长线优化。

对推荐引擎来说，一个可信的内容品牌，是能够“遵守平台规则、持续产出特定垂类、内容质量稳定”的品牌，这几点要求分别对应了更新频率、内容分布和内容受众的情况。

我们在处理作弊案例的时候，经常能够看到各种试图利用平台漏洞来获取流量的行为。比如堆砌关键字，在标题和正文中大量引用各种名人、知名企业等；或是在视频的摘要中写一些无关的词组以误导推荐引擎对内容的理解；抑或购买刷量服务，对内容刷量、对账号刷粉。

作弊与反作弊一定是此消彼长的过程，作弊是能够获得短期红利的。但是，每一次的反作弊升级都能够破解之前的作弊行为，有作弊行为的账号轻则被扣分、禁言，重则被直接封号。作弊只能获得一时的利益，一旦丧失了平台的信任，后续将承受的是断崖式的流量下跌。

举一个脸谱网的例子，脸谱网最近专门打压了内容封面图采用视频截图的情况。这些内容的封面图是一个视频播放器的界面，在图片上展示播放按钮，从而让用户误以为这是一则视频，点击则跳转进入内容。这种作弊方法的脑回路之清新也确实令人叹为观止，但是一旦被发现，曾经发布过这类内容的媒体难免会被秋后算账，遭到脸谱网的降权。

奔着长线目标去经营，做白帽而非黑帽优化，才是自媒体更优的选择。

让用户愿意点击你的内容

内容的包装需要自媒体结合平台特点来进行对应的优化。以今日头条的平台为例：三段式标题、三幅封面图、使用双标题可以说是必须要使用的功能。

三段式标题，是将标题分为三个部分，这样能够承载更多信息量，使用户更容易理解，而层层递进的方式也让表述更为清晰。比如，“北京第一网红餐厅，不用服务员盘子满天飞，

出餐到用餐仅17秒”，“宁波最后的书场，6块钱听一下午，火爆数十年如今只剩80位听众”，给读者留有悬念、制造冲突或是以问句结尾等方式，通常能够吸引更多用户的点击。

三张封面图，是基于App内容展示样式所做出的必然选择。对内容来说，有三图、无图、一图三种样式。三图的样式无论在展示的高度还是在视觉冲击力方面都毫无疑问地具有优势，能够更好地吸引用户点击。结合已有的实验数据来看，对于同一篇内容，使用三图方式的点击率是略高于单图、显著高于无图样式的。

双标题其实是个续命的脑洞功能。推荐引擎依赖冷启动的反馈来决定一篇内容可能达到的分发量级。双标题，给了一篇内容两次冷启动的机会，降低了好作品被埋没的概率。有些自媒体在创作过程中，已经在利用自己的社群或是自有网站进行标题的AB实验。而今日头条，则相当于将这样一个赛马的机会落实在平台层，让好内容能够跑得更远。

近来，头条号还推出了优化助手的功能，旨在最大化地透明推荐参考因素，建议有权限的自媒体使用。

与平台保持积极良好的沟通

上面说的三点都是内容运营层面的。作为内容创作者，同样需要有平台运营的意识。平台规则是死的，但是平台的运营

者是活的，与平台运营者保持频繁的互动有助于及时获悉平台动态，以便更早地参与平台迭代。

在平台的运行过程中，每一次迭代是需要听到创作者反馈的，在各种功能上线前，也都会在核心创作者群试用，并根据反馈来决定功能的去留以及进一步改进的方向。能够有资格加入各平台官方构建的创作者群，其实是对创作者平台运营能力的最低考验。

只有沟通得多了，才能建立更好的互信基础。在碰到如功能使用、内容分发的问题时，及时向平台反馈、帮平台查遗补缺，不仅可以使自己的问题得到更好的解决，也可以在第一时间了解平台的动向。

比如：在某一新版机器消重策略上线后，我们抽样得到的消重效果其实有显著的改进；但是，直到接到特定自媒体反馈后我们才意识到，新版消重策略对于如围棋棋谱、足彩分析等大量引用相似图片的内容，其误消重是非常严重的。有反馈的作者，在策略没有改进之前就已经得到了豁免权限，避免被误伤，并通过自己的内容给系统提供了训练样例，从而让消重策略对自身所在领域的适应度更好。

粉丝运营：新时代的新问题

对自媒体来说，粉丝群体的多寡意味着品牌影响力的大小，并会进一步影响变现能力的大小。无论是打广告还是卖商品，粉丝群体都能够较普通用户有更高的转化率。因此，对自媒体来说，粉丝运营的重要性不言而喻。

如何运营粉丝与粉丝互动，各个自媒体是行家，我就不在此班门弄斧了。但随着全平台分发时代的到来，随着分发方式越来越多元化，我们或许需要回答两个新问题。

第一，推荐平台是否需要粉丝？

第二，是否需要把粉丝导流到微博或者微信上来？

对于第一个问题，推荐平台当然需要粉丝。一方面，推荐平台的运营者需要粉丝数作为参考；另一方面，推荐算法也需要关注粉丝行为作为推荐的参考因子。

站在平台的角度看，什么样的自媒体是有价值的、可信赖的？阅读的数据表现固然能够说明一部分问题，但粉丝的认可或许才能够证明自媒体的长线价值。我们在数据分析中，常常能够看到某些账号通过抖机灵、蹭热点的方式，使单篇内容获得大量的阅读量，但这些账号的内容稀缺性和持续阅读价值是低的，从而其内容阅读量波动巨大，转粉率也往往低于平均水平。

此外，平台方的终极利益诉求在于用户的长线留存，订阅

关系与粉丝互动是服务于这一长线目的的。当用户对某个账号产生认知和依赖的时候，他会刻意地“追更”，从而促进了用户的活跃。平台方对订阅关系的诉求会表现在推荐算法的倾向性中：在推荐过程中，粉丝的阅读表现、账号或内容的转粉情况都会影响内容的推荐效果。

在推荐平台上，粉丝对阅读量的贡献虽然不是决定因素，但一定是显著因素。从统计的角度看，十万粉丝账号的阅读量中位数一定是高于一万粉丝账号的。

对于第二个问题，如果是在2016年年初发问，我会毫不犹豫地回答“是”。彼时，微信和微博的基建完善性，使其成了最好的粉丝沉淀和变现平台。多平台分发的账号往往在自己文中的图片里、文末的介绍里留下小尾巴，引导用户关注自己的微信公众号。

但随着时间的推移，局面正在逐渐发生变化，这个问题或许可以有新的答案。

一方面，微信公众号本身存在信息过载的问题，用户连主动打开公众号列表的动力都下降了，更别提按图索骥完成“打开微信、输入公众号ID、关注特定账号”这一系列操作了。从我个人和身边朋友的例子看，从其他平台往微信、微博上导流的效率，正在逐步降低。

另一方面，站在平台方的角度，其越来越不愿意自媒体将自己平台上的用户导流到其他平台。在平台方完善了自媒体变现所需的基础建设后（如粉丝管理功能、外链扩展功能、用户身份识别的软件开发工具包接入等），平台对待导流的态度也自然趋于严格。

以UC大鱼号为例，官方制定了订阅号推广规范：给作者提供一个特定的推广微信公众号的位置，但是正文中（包括文字和图片）将不得出现任何导流内容，否则将不予推荐。如果被定义为恶意导流账号，还会受到下线内容甚至封禁账号的处罚。官方引导自媒体经营UC大鱼号粉丝的意图，已经表露无疑。

在当下，自媒体是否选择把内容从平台A导流到微信或微博，其实是个生意经：

- 导流收益 = 平台A阅读量 × 平台A到微信平台的转粉率 × 微信的粉丝变现效率
- 不导流收益 = 平台A阅读量 × 平台A的转粉率 × 平台A的粉丝变现效率

哪怕平台A的粉丝变现效率只有微信平台的1/10，但如果从平台A到微信平台的转粉率因为用户习惯或是平台限制的原因只有1/100的话，那么不导流的收益仍然是导流的10倍，在

平台A上踏踏实实地运营粉丝就是划算的。

现在，已有很多完善的第三方工具可以承载自媒体的粉丝互动和变现需求。比如，我认识的很多自媒体都选用了有赞平台作为最终的粉丝服务承载平台。无论是微博、微信，还是其他自媒体平台，都构成了流量运营的入口，最终可以通过外链的方式收敛到同一平台，这使得对接多平台的成本相对可控。

既然接受了全平台内容运营的现实，有精力的自媒体的运营策略或许可以从“导流到微信、微博进行粉丝运营”转变为“全平台的粉丝运营”。只有这样，我们才能够将粉丝增长、粉丝活跃、粉丝变现做成一件全面的事情，针对不同渠道进行专项的优化，从而实现收益的最大化。

全平台运营：从小作坊到MCN

首先，我还是坚持一贯的观点：

创作首先是内向性的，是表达给自己、不在乎受众反馈的，阳春白雪也好、下里巴人也罢，只要创作者达到自己的倾诉目的即可。而当给创作冠以自媒体的名头之后，它就演变成了一门生意，对内需要对团队负责，对外需要面对潜在受众和投资人、广告主。

从成长的角度出发，自媒体或许也会经历从小作坊到MCN（多频道网络）或MPN（多平台网络）的进化之路。

小作坊：大浪淘沙

同很多大号相谈后，我们发现：很多自媒体的诞生仅出于记录感和分享欲。与其将这种类型的内容创作称之为自媒体，不如将其称之为新时代的个人博客。平台和内容载体的兴起会给自媒体带来一定的窗口期流量红利，幸运者往往能够迅速积累起第一批万量级的粉丝，从而开始重新审视自己的内容创作过程：一夜间，发现自己从生活中的路人变成网络上小有名气的小号，每个月也能够靠平台分成挣五六千元甚至上万元了。试问，你会怎么做？

事实上，我接触过的大多数朋友都是在这个时间点上按捺不住，开始跃跃欲试了：从个体的业余创作转向单人全职创作，或是雇用个把人租间办公室进行创作。原本闲云野鹤般的行文也转为更为刻意的内容组织、全平台分发，就连案头的卡夫卡也换成了《社群经济指南》。

咦，自媒体，还真是个全民创业的好项目？

然而，很多人是不适合做自媒体的，就像是杂志的撰稿人不一定真的适合办杂志一样。靠内容积累声名的他们，其能力结构也往往是以低频创作为主，缺乏复合经验，无法很好地把

手艺变成一门生意。其原因主要有以下两方面：

一方面，为了追求曝光量，自媒体的创作需要保证一定的频率。如何在高频的内容创作下保证质量，这本身就是对小作坊式创作的一种考验。如果为了数量而在质量方面妥协，在粉丝数、粉丝互动的层面又会进一步降低。

另外一方面，自媒体创作不能免俗地要涉及钱和安全感。从原本工作中抽离出来的自由人，往往缺乏组织曾带来的安全感，会更在意内容平台带来的收入是否可以超过原有工资、是否能够覆盖成本。那些建立了小作坊的自媒体人，更是睁眼就得琢磨工资的问题。得失心重了，在内容或粉丝运营方面往往更容易急功近利，甚至有些人开始钻研获取流量的歪路子。

对于打算从业余创作转向全职运营的自媒体人，我通常建议他们再多点耐心，至少先等到创作能力稳定和多平台分成收益稳定之后，再迈出全职的那一步。这样的话，哪怕小作坊没有长成托拉斯 [1] ，至少也能维系自己的生活状态。我们常见的以小作坊形式经营得比较好的自媒体，其运营者大多是曾经的媒体人，他们有更职业的内容产出能力、熟悉一些基础的分发套路，即使自媒体创业失败还是能够回归传统媒体的。

制定了养活自己这样一个最基本的目标后，小作坊式的自媒体也算正式杀入了自媒体圈的竞争，从而面对更多来自职业化团队的竞争。

自媒体变现是一个被验证了的市场，有越来越多嗅到其中商业机会的人投身其中。从视频自媒体来看，基本可以分为三种：第一种，已有的图文自媒体升级，基于自己团队的内容内核进行视频化的外部包装，有天然的粉丝基数和品牌效应；第二种，电视台、文化公司的正规军离职创业，从长视频切入短视频，内容创作能力更专业，更容易拿到天使投资；第三种，已经活跃在视频网站的视频制作者，找到了新的分发平台，更容易得到新平台的扶持。但无论哪种，在战斗力上都比小作坊有优势。

“从平常心到功利心。”一个美食自媒体人这么跟我陈述他的心路历程，“就像是个烘焙爱好者突然开了家面包店，挫折比喜悦多，欲望比能力大。”他是在自己的美食自拍集均有5万播放量的时候，选择全职做美食自媒体的。在全职之后，他投入了更多的精力和成本进行内容创作，视频的播放量一度冲击到了10万次以上。但是随着大量美食自媒体转向视频创作，内容供给的大繁荣带来了流量的分流，其视频播放量也发生了回落。好在，由于他的自媒体创立较早，跟平台保持了良性的互动关系，在某次平台组织的优秀创作者峰会上，他认识了投资人与MCN公司，小作坊一点点地要长成大公司了。

以小作坊的形式作战的自媒体人，度过红利期后会逐渐呈两极分化的状态。胜者签约MCN或平台，拿投资招兵买马；败者黯然离场，甚至赌气的称“再也不会提起笔了”。

大浪淘沙，几家欢喜几家忧……

MCN：矩阵思路

MCN是个舶来词，全称为“Multi-Channel Network”。援引维基百科的定义，MCN给视频自媒体（YouTube称之为频道，即Channel）提供如产品、研发、推广（cross-promotion）、数字版权服务、商业化变现等方面的支持，通过与自媒体分成来获得营收。

怎么样，乍听下来是不是像一个经纪公司的活计？

的确，MCN通过提供各种代运营和功能对接服务，在平台和创作者之间构筑起一个中间层。

对平台方而言，同MCN合作的最大收益在于沟通与管理成本的降低。在1万个独立创作者和100个MCN机构之间，平台显然会倾向于同后者沟通，大量烦琐的运营指导、质量监测工作相当于转嫁给MCN来完成。MCN出于自身利益考虑，通常都与平台保持良好的沟通渠道，能够及时获悉平台的运营走向，配合平台的运营方向。援引相关数据，YouTube上100家最大的MCN已经占了平台四成左右的流量。

对自媒体而言，同MCN合作能够在分发层（流量）和变现层（收入）获得收益。以YouTube上的MCN工作方式为例：

- 数据分析：借助YouTube的数据分析支持，MCN可以开发出相应的系统对视频内容的受欢迎程度进行分析，对目标人群进行刻画，辅助创作者更好地创作受欢迎的内容。图11–1即为MCN利用YouTube数据接口产出的分析。

- 流量优化：深刻理解平台规则和运营特点的MCN，有十多个不同的渠道可以在YouTube上获取流量。比如：结合目标人群特点进行搜索的SEO优化，在旗下多个节目间利用annotations（注解）进行流量流转，进行数字化广告投放，等等。

- 内容变现：在YouTube平台内部，MCN有机会获取更高的广告分成，并帮助视频创作者获得官方的原创扶持资金；在YouTube平台之外，还可以帮助创作者承接广告业务、同电视台合作进行短剧制作和IP衍生品服务等。

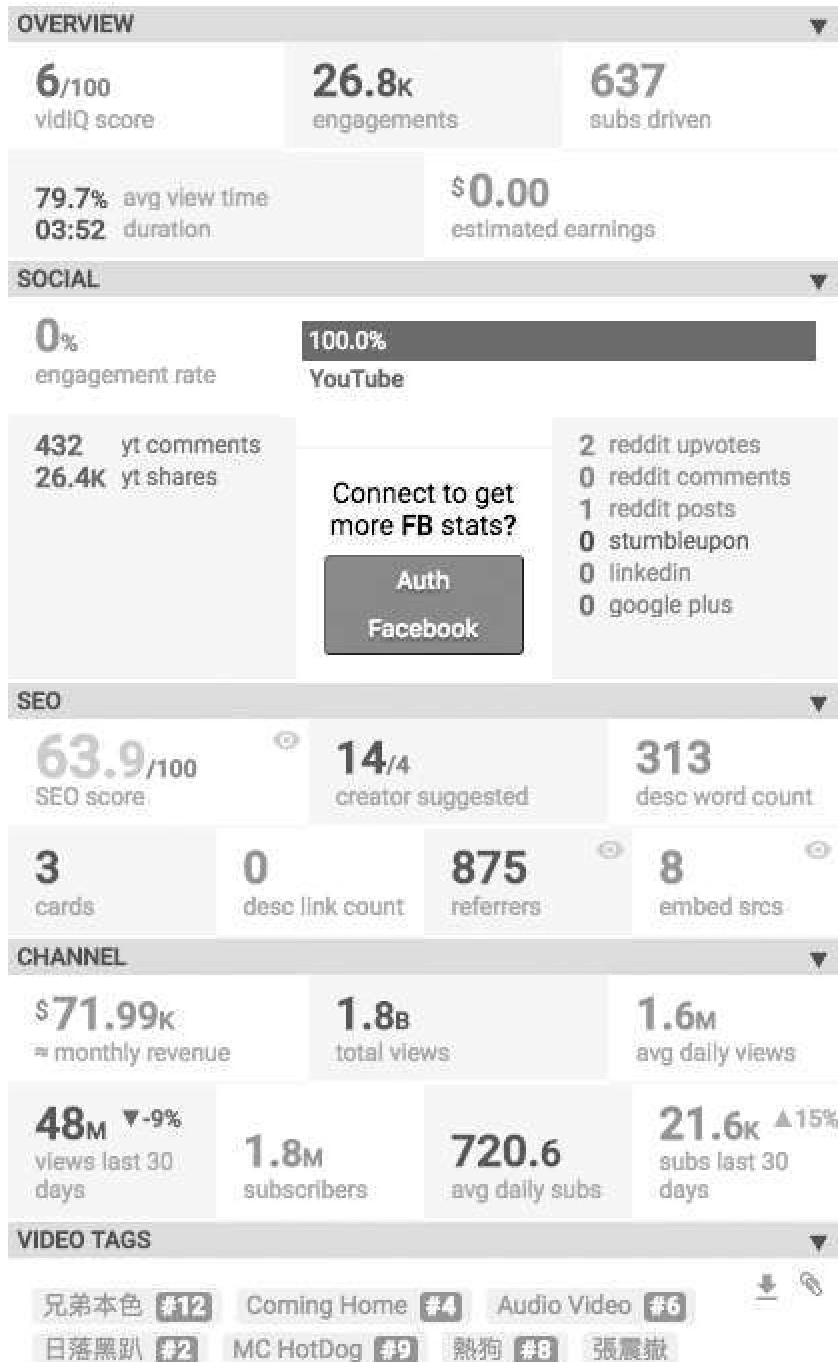


图11–1 利用YouTube数据接口产出的分析

早期的MCN类似于经纪公司，通常更倾向于同头部IP合

作。通过精选一批有影响力或有潜力的视频创作者，对其内容进行深度包装、广告合作来获得低频大额的广告及衍生品收入。这一时期的MCN公司，更多的是运营人力密集型产业。

而随着YouTube平台的运营规则趋于稳定，数据接口和投放接口不断开放，依附其上的MCN公司也逐步完成了从运营人力密集型产业向研发人力密集型产业的转变。自动化程度的提高大大降低了服务成本，使MCN的服务能够普惠到更多作者，从最初签约门槛数十万量级订阅一路降低到了万量级甚至更低。

在同Maker Studio（YouTube的大型视频提供商之一）沟通的过程中，我不止一次地听到他们谈论基于YouTube开放平台所做的数据分析和创作管理，不仅能够帮助他们更快速地发掘出有潜力的视频作者，发现搜索趋势指导创作，还能够借助机器投放批量完成账号与账号间的互推、不同广告位置的露出平台等。

在国外，典型的MCN公司有：

- Awsomeness TV，YouTube上青少年订阅量第一的频道，拥有460万订阅用户，月均千万浏览量。2013年，Awsomeness TV被梦工厂以3300万美元收购。
- Maker Studio，共辖7万个频道，用户涵盖方方面

面，月均数十亿次播放。2014年，该公司被迪士尼以5亿美元收购。

MCN的概念还没有讲清楚，一个新的概念又来了，那就是MPN（Multi-Platform Network），即多平台网络。从MCN到MPN的迁移，实际代表了分发渠道的变革：对美国市场而言，自媒体视频流量不再是YouTube一家独大了。

当脸谱网宣布其视频播放量超越YouTube时，一直深度依赖YouTube的MCN迎来了新的变局，流量的出口从未如此繁荣：YouTube、Vine、Instagram等社交性平台，Netflix、Hulu等SVOD（付费点播）平台，以及传统的HBO、AMC、NBC（美国全国广播公司）等OTT（互联网电视台）平台。如牧民逐水草而居，寻求更多元流量的MCN必然会对接不同的平台方，试图获取更多的利益。

在国内，MCN的概念其实也一直在业内传播，搜狐视频算是“视频MCN”概念最早的践行者，创造了“视频出品人联盟”的概念。从其官方描述上来看，与MCN无异：“拥有大量自媒体出品人资源的机构直接与搜狐视频合作；由其旗下签约的出品人提供视频内容，联盟盟主管理成员收益归属。全想全星、壹酷文化都属于此类联盟。”

2016年，短视频大爆发之后，MCN这个概念再次被各平

台方浓墨重彩地提及，“与MCN合作共赢”成了见诸各大平台公关稿的高频字眼。事实上，由于国内的分发平台比较分散，依赖单一平台难以获取足够流量，国内的MCN公司比国外同行更早地完成了从单平台到多平台MPN的迁移，开始运营微博、微信、美拍等平台。为了方便理解，我们在措辞上仍然使用MCN一词。

进一步地，如果我们关注MCN在自媒体服务、平台对接和商务变现合作上的作用的话，可以不局限在视频领域。在各种直播应用上，工会可以被视作MCN；在微博上，鼓山文化这样的公司亦可被看作MCN。它们都在利用平台规则放大所辖账号的影响规模（通过连麦、拉票、转发、广告投放等运营方式），与广告主对接实现商业化变现，也在与平台的不断磨合中一次次调整自己的发展方向。

以微博为例，其将MCN的广义进一步拓展：“指有能力服务和管理某领域中有一定规模的微博账号的机构，内容形式不限于视频，也包括直播、图文等多种形式。”2016年9月，微博对外发声启动MCN管理系统内测，为MCN机构提供成员管理、资源投放、商业变现、数据分析四大功能。同微博进行MCN合作，最大的好处就是可以获得官方提供的资源扶持和政策倾斜，包括账号推荐、内容推荐等资源，以有效帮助机构旗下账号增长粉丝、扩大曝光、提升影响力。援引官方说明：

2016年，微博在短视频领域与超过200家视频MCN机构进行了接入合作，包括魔力TV、蜂群文化、二更视频、苏州大禹、一条等。在不到一年的时间里，这些MCN机构旗下月阅读量大于10万的账号数提升了118%，单账号粉丝量平均增长52%，账号的阅读总量提升了126%，账号的视频播放总量增长超过50倍。

但是，我们需要承认的是，国内各家分发平台对MCN的信息化支持是相对偏弱的，官方提供的数字化投放方式也相对较少。在前些年，视频网站上的自媒体视频内容相当繁荣，但是国内视频网站聚焦影视剧的原因，决定了平台研发运营资源倾斜的方向。这种情况在最近一两年才有所改善，平台试图补充数据分析和流量流转方面的短板，只有让MCN更高效地管理创作者（成员管理）、了解内容与平台受众间的契合度（数据分析）、有更丰富的手段运营流量（资源投放）后，才可能产生MCN做大的空间。

客观条件的限制导致目前国内MCN机构是运营人力密集型的，无法依赖机器提高效率。这就使得国内的MCN在服务对象上的门槛更高，更愿意覆盖并绑定腰部以上账号，用人工把有限的资源调配给所辖账号。

对小作坊式的自媒体来说，要么以独立工作室的形态签约MCN，要么拿投资签品牌自己成为MCN，这都是发展路径上

的必经选项。但是，在做出选择之前，请先想清楚自己想要什么，MCN公司想要什么，只有同时满足双方的预期才是一门宾主尽欢的好生意。

签约MCN公司的自媒体就像是被星探挖掘的街头艺人，可以得到更系统化的培训和包装，迈出成功的第一步了。

通常在合作初期，MCN公司仅会投入标配的资源，帮助散兵作战的自媒体人适应更加正规化的作战方式：无论是对账号的横向、纵向对比，对内容更完善的消费数据分析，还是平台上多账号之间的互推等。在这个过程中，自媒体人应当多学习运营知识，而不是只做甩手掌柜，把内容以外的部分外包出去就不管了，要多了解自己的内容是如何被MCN运营推广的，了解不同平台的特点。毕竟，在多平台运营上，有着多内容品牌的MCN才可能是有实力说话的。

如果双方配合妥当的话，就像是艺能新人崭露头角，自媒体人很快会进入下一个临界点：即是否要与MCN公司进行更紧密的合作。摆在自媒体人面前的将会是一纸协定：答应绑定或接受投资，MCN公司将会投入进一步的资源帮其做大；不答应绑定，那么后续继续保持若即若离的关系。

以新片场旗下的视频品牌“造物集”为例，其品牌的发掘和培养过程宛若一部路人被星探发掘，签约培养，走上人生巅峰的电视剧。造物集的创作者本是新片场社区的一对小夫妻，老

公把老婆做东西的过程拍下来传到社区与大家分享。被新片场发现后，就将其邀请到北京，提供制作资金和推广资源，并最终成立了单独的公司运营这个品牌，目前已经累积了200万粉丝。

如果你更看重自己的独立性，不想签约MCN“经纪公司”怎么办？如同范冰冰所说的那样，不想嫁给豪门就自己成为豪门。以自身为主IP进行引流，选择与自己的目标人群一致但切入点有差异的内容品牌进行扶持，借助自己的内容市场化创作、粉丝与流量运营、平台合作经验来指导新的内容品牌。

以papi酱为例，在其获得高额投资后就成立了以自己为主IP的MCN公司papitube，开始孵化和培养新的内容创作者，签约了Bigger研究所、ACui阿崔、张猫要练嘴皮子、KatAndSid等多个知名账号。

在2017年异军突起的“办公室小野”，其公司洋葱视频也陆续推出了“办公室小作”和“办公室小五”等办公室系列短视频，其MCN矩阵将会延续“办公室”基因，拓展该系列旗下的新IP。

成为MCN，其实更像是一种内容矩阵的升级版。对内容这样的创业产业而言，“人无千日好，花无百日红”，再强大的内容IP都可能有过气的那一天。但是，妥善经营的内容公司可以有比IP更持久的生命力。

MCN公司的生命力不在于能够创造一个单独的爆款IP，而在于能够持续发掘、捧红一批IP。如果非要选择一个标杆的话，非迪士尼莫属。

“面向市场生产内容，面向平台分发内容，面向用户运营品牌”成了很多MCN的经营思路。

典型MCN分析：新片场

新片场是国内由视频创作者社区“V电影”发展而来的一家MCN公司，并于2015年登陆新三板，成为一家上市公司。透过其相关公开演讲和年度报告，我们可以更好地了解MCN这门生意在中国是怎么做的。

在创作者对接上，新片场目前管理着300多个频道，全网累计粉丝破2亿，播放量破60亿。其内容分为自制和签约两部分。对于A类内容，为了增强控制往往以自制或投资控股等方式进行强连接，其余内容则是通过签约的方式弱关联。新片场体系中影响面比较大的短视频品牌，几乎全都是自营品牌，如“造物集”“魔力美食”“小情书”等。



图11–2 新片场内容分类

考虑到运营人力投入和利益分配的问题，这种情况并不难理解：已经坐拥平台方优惠政策的MCN，当下的服务方式尚依赖一定规模的人力投入，无法大规模地进行自动化，故只能将有限的精力投入到部分IP上，通过自制或控股的方式深度绑定视频IP，才能够最大化投入产出比，优化自身的长期利益。

在某次接受36氪采访时，新片场CEO（首席执行官）尹兴良直言不讳道："签约在整个新片场体系内不占很大比重，我们整个业务的重点也不在签约上，大部分内容品牌是我们自己管理的。这些内容品牌都是我们自己从零开始或者重度参与孵化、创作的。"

在平台关系上，由于新片场管理了大量的视频品牌，毫无疑问地成了各大内容分发平台的座上宾。新片场也发挥了与平台沟通充分、对平台定位了解的优势，将内容创作变成为平台定制生产的方式，使不同节目能够更适合不同平台的分发调

性。

与平台的深度合作使其播放数据暴涨。以魔力TV系列子品牌为例，在与微博进行MCN合作后，它旗下月阅读量大于10万的账号规模由最初的18个提升到54个，单账号粉丝量提升了75%，账号的阅读总量提升了350%，账号的视频播放总量增长超过200倍。

罗列一下新片场获得的殊荣，可见其与平台的关系：获2016微博V影响力峰会“十大影响力视频机构”称号，旗下“尖叫制片厂”获微博“十大影响力幽默博主”，“魔力美食”获微博“十大影响力美食博主”，并获得今日头条2016年“年度头条号”，等等。

在收益变现上，除了基础的平台方流量分成外，新片场也在探索如周边衍生、软广植入、内容电商、IP孵化等方式。根据新片场年报，2016年公司在周边产品及衍生品方面实现收入400万元；旗下视频品牌造物集开设的电商店铺，实现年营收五六百万元；与雪碧合作的软广视频“6种方法让雪碧更好喝”在微博上的播放量突破1000万次；双十一期间，与淘宝、天猫共同尝试做内容电商，推出“脑洞研究所”“会动的好货”“礼遇季”等短视频节目，以内容传播带动商品销售等。

此外，新片场需要注意的一个问题是：受限于短视频市场的整体变现情况，MCN目前还没那么赚钱。尽管新片场在短视

频部分的分成收益、广告变现已经相对很高了，但是分析其财报，我们可以发现：真正构成公司大规模收益的是通过网络电影、网络剧从视频网站获得了可观的内容点播分成收入（如图11–3）。这或许也解释了为什么很多短视频创作者都希望向大IP、大荧幕发展：天下熙熙为利来，天下攘攘为利往。

（4）应收账款金额前五名单位情况

单位名称	与本公司关系	金额	账龄	占应收账款总额的比例	坏账准备期末余额
霍尔果斯乐视新生代文化传媒有限公司	客户	31,757,615.30	6个月以内、6个月-1年	71.32	3,805,678.40
北京爱奇艺科技有限公司	客户	10,751,793.50	6个月以内	24.15	
北京快跑小鸡文化传媒有限公司	客户	1,096,021.00	6个月以内	2.46	
浙江天猫信息技术有限公司	客户	450,000.00	6个月以内	1.01	
上海全土豆文化传播有限公司	客户	120,000.00	6个月以内	0.27	
合计		44,175,429.80		99.21	3,805,678.40

图11–3 新片场2016年财报（部分）截图

[1] 托拉斯：垄断组织的高级形式之一。——编者注

本书由“ePUBw.COM”整理，ePUBw.COM 提供最新最全的优质电子书下载！！！

自媒体变现

首先，我们大喊一下口号：一切不以赚钱为目的的商业自媒体，都是耍流氓！

对由兴趣创作升级为商业创作的自媒体而言，营收是一个大问题：要想向前进，势必要离钱近。目前，自媒体从哪些途径获得收益？各种收益的规模和获得收益的典型方式是怎样的？

从我个人的观察来看，自媒体的收益来源可以分为两种：流量变现和内容变现。前者将内容作为流量入口，基于流量进行后续的变现行为；后者将内容作为产品，借助内外部流量来推销内容。在这一章中，我将逐一介绍每种方式。

变现入门：平台分成

流量变现的最初级形式就是依靠平台的分成实现收益，这也是原创实力较弱或者知名度不够的腰部账号的主要收入来源之一。

对于腰部账号，收入覆盖成本是必须考虑的运营问题。如

果先期不考虑采取内容投放或付费运营的方式，单纯依靠内容从平台获取流量，那么主要的成本就是人力资源成本。

以每个编辑月收入10000元计算，再考虑五险一金和办公支出，对于公司来说，每个编辑的月成本在15000元左右。以每月20个工作日计算，那么每日每人需要至少创造出750元的收入，才刚够覆盖其人力成本。

问题就收敛为：这750元，从哪里来？

答案是一个简单的公式：分成收益 ＝ 阅读量 ×CPM

根据发展阶段的不同，各平台对于千次阅读CPM的广告分成也不同，通常在0.5~5元的范围内波动。对于发展初期的平台，由于需要快速吸引作者入驻、实现内容覆盖和作者覆盖的目标，故通常采用平台高额补贴的方式进行刺激，作者能够获得的CPM通常较高，存在一定的溢价情况。

对于成熟阶段的平台，内部的流量分布和分成机制趋于稳定，外部的用户群体和社会知名度也达到一定量级，作者能够获得的CPM较为稳定，但是整体收入得益于流量规模，也相对稳定。

除了由于平台发展阶段的不同导致不同平台的分成系数不同之外，我所了解的几家平台对待不同的创作者也都存在不同的分成梯度：被平台识别为优质的自媒体，往往能够获得更高

的CPM。

一如当年的“百团大战”让店家和用户都得到了补贴，考虑到当前的内容平台呈“战国之态”，有吸引力的平台分成策略预计会持续一段时间。

对优质的原创账号而言，除了分成收入外，还有来自各家平台的激励和扶持奖金。结合各家平台已经公布的计划，单个账号的月激励额度普遍在万元左右。

- 头条号：千人万元计划、礼遇计划（在原有收益基础上，头条号平台综合内容质量、原创度和数据表现，每月向100名创作者提供1万元奖励）
- 大鱼号：量子计划、大鱼计划（每期评选出超过1000名创作者获得万元奖金）
- 百家号：“百＋”计划（上榜作者有机会获得最高一万元的创作奖金）
- 企鹅号：芒种计划（100%的广告分成加原创补贴）

回归最开始的数学题，750元的日收入从哪里来？如果仅依赖平台分成，以CPM为1元来计算的话，那么每日需要获得75万次的全平台阅读量，单平台要保持十万次以上的规模。考虑到原创内容的生产速度，十万次以上的阅读量对于腰部账号

来说委实不低，这也是为什么我不建议兴趣写作者贸然转向全职的原因。

但目前的局面对创作成本较低的做号者来说确实存在盈利空间。如同外卖业的兴起催生了一大批灰色外卖食肆一样，内容平台的兴起同样催生了一批以伪原创和搬运为主的做号者，我将在下面进行介绍。

内容分发的江湖客：做号者

有人的地方，就有江湖。

更为准确地说，有利益的地方，就有江湖。

在内容分发领域，我见识了各种各样的内容创作者。因缘际会，我也经朋友介绍认识了这样一个靠平台分成年入百万的做号者，让我们称他为老A。相信很多内容创作者会对做号者的做法不齿，但老A在洞悉平台运营规则和内容包装上的投入，是有一定参考和借鉴意义的。

存在的或许就是合理的，让我们走进和老A的对话。

大雨，老A急匆匆地来，一落座就开始大倒苦水：

“现在生意不好做了，各家平台的监管也越来越严了。你们最近上的策略，封了我手里一百多个号啊。”

“多少？”我生怕自己听错了数量级。

“一百多个。”他重复道。

在惊愕中开场，他开启了我对做号者江湖的认知：从账号的批量注册到模式化的内容生产，再到平台分成和自营广告变现，做号者已经形成了完整的产业链条。

第一步：积攒账号

要想通过平台分成获利，首先需要有大批量的账号，通过多个账号的分成累积最后获得相对可观的收入。目前，内容平台对于账号注册的基础配置通常为：手持身份证验证 ＋ 手机号码验证。

“手持身份证验证？太原始了，基本上通过‘在线搜索 ＋ PS（图像处理）’的方式就可以绕过去。真正的大杀器是批量处理短信和银行卡。”老A不屑地说。

对于手机号码部分，我原本的认知如图12–1所示的机器，在一些测试平台也有广泛的应用。

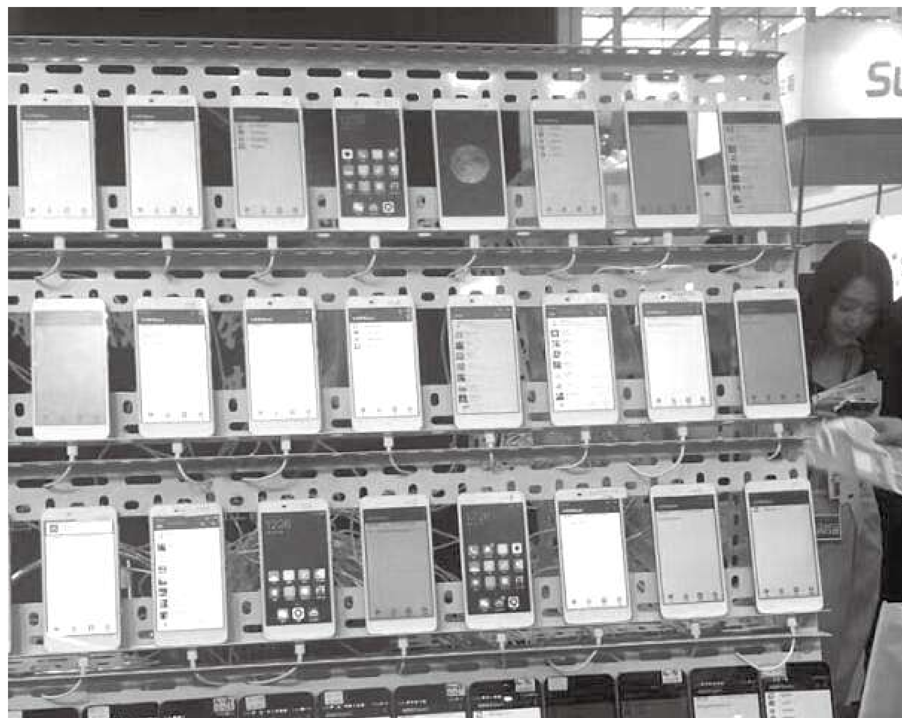


图12–1 多手机并行测试机架

孰料，真正让老A这样的做号者如虎添翼的是如图12–2所示的设备：猫池（ModemPool）。猫池是一种可以同时接入大量SIM卡（用户身份识别卡），模拟手机进行短信收发、语音拨打、流量上网的设备，在银行业、邮局、交易所券商等场景下有广泛的应用。图12–2中展示的就是64口的猫池，单机能够同时接入64张SIM卡，通过软件处理短信验证码。



图12–2 64口的猫池

“等等，手机SIM卡不是已经实名制了吗？为什么还有这么多SIM卡流出呢？”我追问道。

“喏。”老A指指路边的共享单车，“我们用的是那玩意儿。”

原来，这些用于接码的卡并非我们常用的手机SIM卡，而是我们提供给企业用于智能交通、物流管控的专项卡：物联网卡。我们生活中最为熟悉的共享单车摩拜与ofo，其用于定位和解锁的智能锁，就是通过物联网卡与服务端进行通信的。

与普通的SIM卡相比，物联网卡对于做号者的吸引之处在于：

• 零月租或者低月租，大多按照使用的流量进行付费。在某销售物联网卡的代理公司网站上，我查询到的价格是每100M 8元。

• 凭公司营业执照办理，可以一次性大批量办理。某些代理公司因追求业绩而疏于管理，可以在购买者没有营业执照的情况下，代办和代发物联网卡。

• 有语音卡和短信卡两种，前者可以双向接打电话和收发短信，而后者只能接收短信、价格更低廉。

有了这批物联网卡，大多数平台的注册（短信二次验证）机制都可以稳定地绕过了。与此同时，这些物联网卡不仅被应用于内容平台，也可以一卡多用，应用于各种灰产场景。例如，应用于互联网金融、电商等平台赚取利差，应用于刷量、刷粉等领域不正当获利。

由于大多数平台在发展初期的审核识别机制都不甚完善，而后期完善后也倾向于过往不究，所以做号者往往会关注一些新平台的动向，在平台开放注册伊始就大批量囤积账号。

为了对账号的真实性加强控制，平台也逐步增强了对账号的验证，如增加体现环节的校验，要求银行卡与身份证保持一致。

然而，这只是提升了做号者的成本，无法真正杜绝做号现

象。老A向我展示了如图12–3所示的银行卡、身份证、网银U盾和手机卡，行话称为“四件套”，行情价在千元左右。考虑到目前有四五家比较大的提供分成的内容平台，只要单家平台月收入超过200元，一个月就能够收回投入成本。



图12–3 “四件套”示例

“只要用心运营，还是有赚头的。”老A信心满满地说道。

第二步：生产内容

有了大量的账号，还需要有可持续的内容生产能力并通过内容获取流量才能够换取分成收入。我向老A提及自己了解到的同义词替换、中英翻译、多篇文章截取段落拼凑成一篇文章等攒文手段，他一脸瞧不上的表情：“那些都是低端的，我玩的是高端的。”

他推了推眼镜，说道："我可是新闻系毕业的，攒内容，我们是专业的。"

在老A的团队里，批量内容生产分为两个步骤：背景调研和快速洗稿。

背景调研，是老A投入最大精力做的事情，主要用来确定平台内容创作的大方向，分为全网热点和平台特点两部分。这一部分，也是我觉得值得很多中小原创自媒体参考借鉴的。

全网热点用来洞悉自身领域内部的用户需求，确定账号基调。老A通过"360趋势"来确定每个行业中的热门关键词和热门主题。以健身账号为例，通过搜索"健身"可以得知，与之高度关联的有"中老年健身与气功""囚徒健身""keep（一款具有社交属性的健身工具类产品）""八段锦"等关键字，这些关键字共同确定了这个健身账号后续发文的基调。

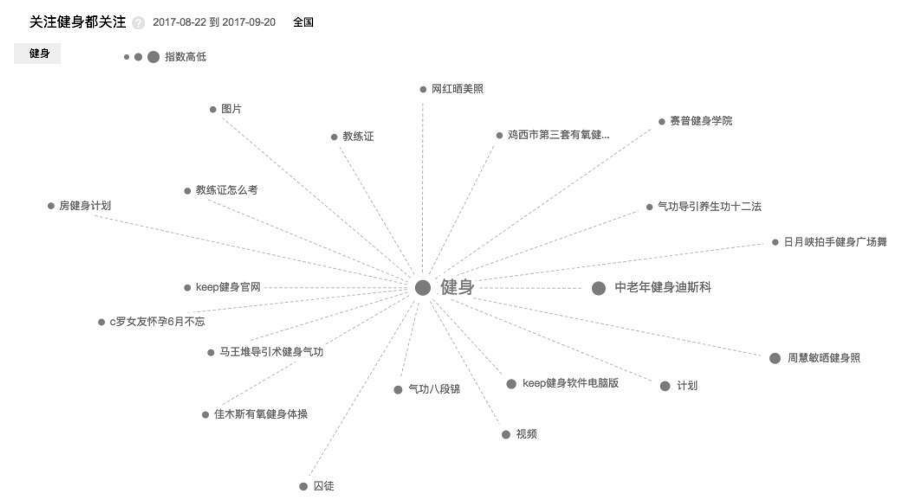


图12–4 “健身”搜索结果截图

“只要所有内容围绕这些关键字开展，就不会太跑偏。”他说道，“我们这也算科学选题了。”

在选题确定之后，老A下一步确定的是平台的调性，看主题下的哪些文章适合这个平台。具体来说，就是在对应平台上搜索主题下的热门关键字，找到那些高阅读量、高评论量的内容标题进行拆分，以分析平台用户或平台分发机制的具体偏好。以搜索“八段锦”为例，高评论量的标题有：

- 八段锦健身有奇效，练习诀窍你掌握了吗？
- 健身气功八段锦视频教学口令版
- 1分钟学会八段锦

“这些标题的评论量高就说明读者是想学，那所有面向老年人的关于气功的内容都得围绕练习、学习、教学展开。算法分发是要让机器读懂的，标题就得把‘练习’‘口令’这些关键字都带上。”老A讲出了兴致，眼闪着光。

确定了领域热门主题和平台热门关键字之后，就进入了原稿选择和洗稿阶段。“这个简单，谁火选谁。”选择一篇领域内的高热内容作为框架，保持主题不变开始洗稿。关于洗稿的方式，老A总结出了三个“东西”法则：“东拉西扯、东拼西凑、指

东骂西”。

- 东拉西扯：交代背景和关键字。比如要写八段锦，那就得介绍八段锦的背景，管它是野史还是正史，先摘抄几句再说。如果近来某个人物比较火，就捎带一句“某某也时常练习八段锦”。

- 东拼西凑：以原文为模板，在网上搜索与之相近的内容，每篇内容摘录一小段，和原文混在一起构成新的内容。拼凑出来的内容，读起来倒也还算通顺，还能够骗过市面上主要的消重软件。

- 指东骂西：在内容的结尾部分就该制造冲突了。谈论八段锦的内容，要么制造中西医之争，比如西医说八段锦纯属伪科学，没什么用处；要么制造中外之争，比如八段锦被拿去申请世界文化遗产了等。

三个“东西”法则下来，一篇火热的原文就已经被改得面目全非、难辨本来面目了。“剩下的就是配上好封面，起个好标题，看命了。”老A言道，最后又追加一句，“不过，我的命目前看起来好像都还不错。”

“一天产量能有多少？”我问道。

“比较有门槛的领域，一小时能洗四五篇；一般领域的内

容，一天能洗30篇吧；要是娱乐科技的内容，一天我能洗出七八十篇。”

老A不无炫耀地递给我一篇稿件，1.3万的阅读量。我浏览一番，除了觉得内容比较泛泛之外，还真心看不出什么明显的问题。再看用户反馈，指东骂西显然起了煽动情绪的作用，大量用户在评论区开始了争论，这或许正中老A的下怀。

第三步：变现探索

我默默地给老A算了笔账：

按照一个人一天30篇，每篇5000次的阅读量来计算，一个人一天的流量就有15万。同时分发在四个平台上，一个人一天能够有60万的阅读量。即使按照最低的CPM 0.5元计算，一个人一天也能够带来300元的收入，全月带来6000元营收。按照每人月工资4000元来计算，一个人能给老A带来的收入是2000元左右。

“一个人身上能赚2000元，你有多少人啊？”

“纠正一个问题，一个人身上能赚4000元。”老A说道，“我雇的可是清一色的大学生实习，开社会实践证明的那种。”

看着我愕然的脸，老A笑笑：“我的工作室不算什么，才几

十个人。我可是听说一个前辈，生生把这事情办成了公司，几百名员工给他打工，有专门的技术部门负责抓新浪微博热搜，从海外搬视频回来后配字幕呢。”

就像是莆田的仿鞋生意，做号者们生生把内容伪原创做成了一门生意。

“你们除了平台分成，会自己做电商或接广告吗？”我问道。

“广告是要接的，流量大了就有中介找上门了。软广硬广都接，要是阅读量不够，就在淘宝上买点，也算过了一段好日子。不过，最近刷不了量啦。”老A不无惋惜地说，“你们上了这玩意，把刷的量都摘出来了。”



图12–5 今日头条内容阅读量示例

我嘿嘿一笑：“应用内外阅读数拆分这事，是我上的。”

老A愣了一下，摇头苦笑道：“也好，也算我找对人了。你听我讲了这么多做号的事情，估计回去就该专项打击了，下面就该我向你请教了。”

“做号，无非是个吃红利的事情，这个道理我懂。就算你的平台不能百分之百地查到，封掉我几百个账号，对我来说也是伤筋动骨的。”老A缓缓地抽了根烟，说道：“我想正经地做个原创自媒体，拿投资，搞个一条二更出来，多风光。”

“这是条路啊，你刚才说的过程除了拼凑内容的部分差点

意思，内容主题和平台调性分析都做得蛮不错的，甚至大部分原创自媒体在运营层面都不一定有你做得细致。”我认同地说道。

“可是不挣钱啊，我试过请个像样的编辑，一个月工资上万元，可是他一天写一两篇内容，挣到的钱还不够给他自己发工资呢。你得给我好好讲讲，那些原创大号是怎么过日子的。”

关于原创自媒体、内容与标题运营，我跟老A又聊了许久。我清楚地感知到，赚快钱的老A对于平台规则和流量运营有着近乎本能的敏感。但是，真正的原创内容、原创自媒体就像是种庄稼，“春种夏耕秋方收”，只有品牌效应和粉丝积累到某个临界点的时候才会上一个台阶，凭借自身的品牌溢价承接高价广告。

“你真的有耐性熬那么久，去做原创自媒体，来换一个可能性吗？”临别时，我问老A。

老A搓搓手：“我再想想，再想想。”

在平台的某次打击过后，老A给我发来了微信：“最近为挣广告费，卖了点保健品，又被你们逮住封了。你们有点儿太严，我得主营别的平台去了。”

我回复：“治理混乱期没那么久的，别的平台早晚也会进

化到这一步。你转型做原创的事情想好了吗？”

但这一次，老A没有再回复我……

广告变现：品牌的溢价

平台的分成，本质上是自媒体将自己的内容流量交给平台方去出售，平台方将获得的广告收益与自媒体按比例结算。如果自媒体能够做出自己的品牌，就有机会独立承接广告。

平台广告分成和自己售卖内容软广的关系，就好比普通橙子和褚橙的销售，在曝光量不变的情况，自有品牌带来的收入溢价不止10倍。也正因为独自承接广告能够显著改善自媒体的收入情况，这也成了大多数原创账号努力的第一个节点目标。

那么，广告主为什么要投自媒体广告而非平台广告呢？

其本质是用经济利益来交换媒体的影响力，实现媒体价值背书的触达。

用一个公式表示自媒体的影响力价值：自媒体的影响力价值＝粉丝影响力强度×粉丝数量。自媒体定位越垂直，其对个体用户的影响力强度就越大；自媒体粉丝量越多，其影响力的覆盖面就越广。在自媒体发展的前期，粉丝影响力强度的意义甚至要大于粉丝数量的意义。毕竟，在注重转化的时代，深度

影响一个人的意义要比泛泛的十次曝光有意义得多。

某些垂类的账号天生具有承载广告的优势，这一点或许值得想要入行的自媒体人关注。比如，面向有付费能力的中产阶层女性的美妆时尚账号，面向亲子教育、紧紧抓住新手爸妈痛点的母婴亲子账号。与历史类、泛娱乐类内容相比，这类内容本身离商品更近，也就拥有更多潜在的大众消费品类广告主。

除了常见的以服务普通用户为主的自媒体外，还有一些立足于垂直行业的自媒体，比如餐饮老板内参、母婴行业观察，由于面向的是经营者，同样会受到2B（对商家，泛指企业）服务企业的青睐。餐饮老板内参在2016年所完成的融资，领投方之一就是深耕餐饮市场的新美大集团。

随着流量分发方式的改变，自媒体广告承接与广告的表现方式也在逐步发生着改变。

在以粉丝分发为主的流量分配方式下，广告的承接方式简单粗暴，自媒体直接将自己的曝光位置出让给广告主进行硬广推广，如微博转发，微信公众号出让头条、二条位置，如图12–6所示。

在效果监控上，投放前通常能够基于账号的过往粉丝阅读量估算出一个基础广告触达量。如果广告够软，阅读性、互动性够强，就能进一步收获更多的转发阅读量。

由于出卖位置填充硬广的方式沟通成本低、操作简单，因此被更多的广告中介所承载，将广告需求分包给相对较小的自媒体账号。而当自媒体做大做强以后，也会更看中自己的品牌价值，对广告主更挑剔，并进一步介入广告内容的制作中。

在推荐分发主导的平台上，自媒体需要重新审视广告触达量的问题：在内容展示量与粉丝量没有强关联的情况下，怎样才能满足对广告主的流量承诺呢？阅读完成率低的硬广，显然无法适应推荐分发的生态。



图12–6 微信公众号出让头条、二条位置截图

如果说在粉丝分发平台上，广告内容化还属于可选项的话，那么在推荐分发平台上，广告内容化就成了必选项。自媒体只有将广告处理得更加内容化、延续自己账号一贯的内容调性和图文风格，才不至于出现“正常内容十万加，广告内容零分发”的窘境。

广告内容化给我留下深刻印象的，是二更在2016年所承接的牛仔裤品牌Lee的广告。2016年8月，二更同Lee合作了这样一则视频：



图12–7 二更与Lee合作视频截图

视频中，几对情侣参加一则测试，要求男友不看脸，仅通过臀部和腿部来猜测哪一位是自己的女友。几次测试下来，由于女友换了裤子导致多位男生都没有找对。问题来了：为什么换了裤子，就能穿出远胜平日的、更好的腿型和臀型？时间回溯到几位女生在更衣间的环节，谜底揭开，原来她们都换上了Lee牛仔裤。从而，得出广告主题：这一件新衣，连男生都觉得应该买。

视频一上线，3天内就获得全网1300万次播放量，仅在今日头条的平台上，这一则视频就以“女生换上一件新衣服，相恋2年的男友竟然认不出，牵错了手！”为题，获得了310万次播放量。这个案例可以说是内容与广告双赢的典型案例了。

二更COO（首席运营官）皮行早就在某次分享中提道：“内容即广告，广告即内容。用户看完这条内容不会产生反感，反而喜欢看这个广告。二更的营收来源中，超过70%来自商业视频原生广告的定制。”

“广告即内容”，这种内容原生广告的形式或许为自媒体实现广告销售与粉丝体验共赢提供了一条出路。好的广告植入，不仅不会对用户阅读体验造成损害，还会出现粉丝看正文猜广告、与自媒体充分互动的情况。以X博士的一篇微信内容为例，在文末“神转折”引出汽车广告后，粉丝对广告也表现出了极大的善意，在评论区中充分地互动，如图12–8所示。



图12–8 X博士汽车广告与粉丝的充分互动

正是由于原生广告的高用户接受度，它也越来越受到广告主的青睐。援引MediaRadar的报告数据，2017年第一季度，美国原生广告购买量在所有广告形式中增幅最高，同比增幅74%。这对擅长内容创作的自媒体来说，是一个非常正面的信

号。

当然，原生广告的创作过程并不是一蹴而就的，包含了反复锤炼的过程。自媒体与广告主的合作流程通常是：承接广告意向，就广告主需求进行初次创作，就创作内容与广告主反复磨合，通过自己的账号进行多平台发布，汇总多平台的数据表现给广告主做结案。其中，最耗费精力的是就创作内容与广告主反复磨合的过程。账号规模越小，话语权也就越小，返工的次数就会越多。我就不止一次地听到内容创作者对广告主的抱怨，这或许是因为他们还没有体会过乙方角色的原因吧。

当一个自媒体能够独立承接广告后，它也正式从内容平台的投稿者转化成了利用内容平台进行流量分发的媒体角色。达到这一层次的账号，往往一个大单就能够覆盖一个月的成本。个中翘楚如公众号“咪蒙”，网传头条广告价格每则60万元，二条每则35万元更是令一般自媒体望尘莫及。

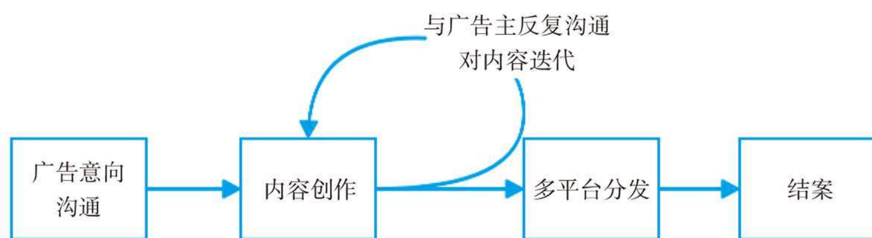


图12–9 自媒体与广告主的合作流程

有了收入的自媒体可以用商业化的方式来运作账号，自身也会进化为平台的广告主，通过广告投放来快速扩大自己在平

台上的影响力和号召力，开启自媒体成长的新篇章。

广告中介：是否是一门好生意

世间总总，大抵逃不过二八定律。对于内容产业也是如此：80%的广告收益，被20%的自媒体占据。这归因于其品牌影响力，头部自媒体有能力对接可以支付溢价的大品牌客户，切走了广告市场最大块的蛋糕。

当然，剩下20%的收益也足够大了，于是就有了如新榜、微播易、天下秀、微梦这样的广告中间商，试图对接20%的广告收益和80%的自媒体。

如前面章节中所提及的，由于目前各分发平台的信息化程度不够，使得中国的MCN公司很难像国外同行那样规模化地给海量创作者提供全面的服务（数据分析、平台运营、数字化投放、广告对接等）。自媒体服务中相对容易标准化的一环——广告对接就被抽离了出来，促成了更容易规模化的广告中间商生意。

广告中间商，做的是信息不对称的中介生意。一边是对预算精打细算的广告主难以找到匹配的自媒体，一边是没有足够知名度或单独招商无门的自媒体，广告中间商将两者对接并从中赚取中介费用，类似网盟的模式。如图12–10，即为微梦对自己模式的描述。



图12–10 微梦模式描述

这门生意看起来四平八稳，实则也有相对明确的边界和天花板。

站在自媒体的角度，自媒体本身是有成长路径的，处于不同阶段的自媒体对广告中介的依赖程度和诉求是不一样的。

图12–11是火星文化CEO李浩对自媒体发展阶段的进阶分析。他将自媒体以估值千万、亿、十亿、百亿元作为区隔分为五个阶层。从图中我们不难看出，规模越小的自媒体越亟需广告中介的信息对接。



图**12–11** 李浩对自媒体发展阶段的分析

千万元估值以下的自媒体，多是集中于内容创作本身的小作坊形态，没有自己独立的商务团队，品牌辨识度相对较弱。在这种情况下，自媒体通过平台赚分成和通过广告中介变现的关系，就像是出租车从受雇于租车公司领工资变成服务于滴滴平台干个体，只是通过广告中介获得的收益比平台分成略高。

大量的小额广告订单和中小自媒体的对接更接近网盟形式，中介平台将广告资源做竞标或分销，只提供信息的集散广场，不会投入过多的专项人力运营服务。我们可以通过查看新榜上的“新榜有赚”频道，来了解这一部分广告资源与流量资源的对接情况。



女频小说推广

阳江市江城区博弈手机店

竞标数：122　预算：10000.00



企业传播何以制胜？16…

引爆点互动科技（北京）有限…

竞标数：85　预算：8000.00



国庆出游，你有一份特…

深圳出行管家网络科技有限公司

竞标数：53　预算：2000.00



全球最贵￥4,448麦片…

厦门鑫宇进出口有限公司

竞标数：42　预算：2000.00

图12–12 “新榜有赚”页面

随着自身的不断成长，千万元估值以上的自媒体运作已经从单纯的个体内容创作行为演化为包含团队内容创作、流量运营、商务拓展等的公司行为。这一阶段的公司具有一定的垂直行业知名度，自然不会仅仅满足于出让位置给广告主，会更多地进行“广告即内容”的广告创作。在这一过程中，信息平台的分发对接已经不能满足需求了，需要更多商务、创意人员的人工介入。

小龙套会成长为大明星，人红了就可能会解约经纪公司创

立自己的工作室。很多上规模的自媒体已经开始谋求相对独立的商业化变现。以MakerStudio的头号网红为例，PewDiePie在2016年与MakerStudio合作推出了自己的MCN——Revelmode，从事付费内容、游戏研发、慈善项目、商业合作以及周边产品业务。

就像各大门户网站，其核心广告位一定是由自己销售，只有边边角角的流量入口才会交给网盟来变现一样，面对有一定规模的自媒体，广告中间商并没有足够强的议价能力，只能成为分销渠道或信息渠道之一。比如，火星文化是为“军武次位面”提供运营和变现服务的，但军武自己也在进行游戏广告、自营电商的变现探索。

规模相当的自媒体之间进行合作还可能会出现飞单的情况。在首次合作顺利的情况下，如何才能保证广告主与自媒体后续依然通过中介平台进行交易呢？

自媒体行业兴起的时间尚短，大多数团队为偏创作的能力结构。在这种背景下，广告中介公司在自媒体变现环节中起到的作用是巨大的。微播易CEO徐扬曾介绍，超过七成自媒体大咖的第一个广告订单来自微播易，40%的明星在社交平台上接下的第一个广告订单也来自微播易。当然，接下来值得发问的是：后续的广告订单是否依然由微播易承载呢？

从自媒体资源的角度看，除非以独代的模式签下某些自媒

体，否则，广告中间商很难稳定地把持高质量和有影响力的自媒体资源。

说完自媒体的角度，再来说说平台的角度。

自媒体宣发的边界是平台，广告中介需要面对的一个大问题是：如果平台也来插手这门生意的话，怎么办？

在这个问题上，微博是最典型的案例。正如你所知的那样，在2012年之前，尽管微博的用户活跃度和社会影响力极高，但是平台方并没有从中赚到多少钱，反而是一帮大号或者草根大号靠发软广、硬广挣了钱。彼时，微播易一度与上千微博大号签订有广告投放协议，赚得盆满钵满。

这一局势的扭转始于微博对中间商天下秀（IMS自媒体商业集团）的投资。天下秀联合微博推出了官方“微任务”平台，直接在中小广告主与大号之间建立了一道桥梁。自此，天下一统，大量微博平台上的广告分发和对接业务逐步收敛到“微任务”上。

和微博相比，微信官方在这方面一直很克制，直到2016年10月才推出了互选广告平台。也正是由于它的克制，才使得市面上的微信广告中间商能够快速发展壮大。以天下秀为例，其除了是微博平台最大的广告中间商之外，同时也以“平台撮合＋人工接单”的方式为大量广告主提供与微信公众号的对接业

务。

相对于独立广告中间商，平台从事广告中介业务，可以说存在碾压性的优势：在数据层面，它掌握着最真实、可信的自媒体分发数据和用户画像，只有官方才知道广告内容超过十万次的阅读量中哪些是真的，哪些是刷的；在流量层面，又直接掌握着平台内各种流量分发渠道。网上流传这样一种说法：如果不通过某平台自己的广告分发平台下单的话，那么自媒体在该平台的广告分发是很难有大流量的。

在数据与流量的双重优势下，当平台决定下水的时候，广告中间商的冬天，还会远吗？

何以过冬？投靠平台，垄断创作端，拥抱未知。

投靠平台：以天下秀为例，能获得平台的注资是最佳策略。背靠大树好乘凉，只要平台不倒，就能利用足够的信息优势和规则优势傲视群雄。

垄断创作端：除了自媒体官方之外，广告中间商可能是最了解自媒体数据的一群人了。利用自己的数据优势去发现早期自媒体，在其尚未出头之时便深度绑定变成长期发展最好的选择，这跟MCN的选择会比较类似。比如，微播易就自营和投资了大量微信公众号，火星文化也先后投资了青藤文化、雯思东方等账号。

拥抱未知：和自媒体内容蹭热点一样，广告中间商也需要持续地追热点。

- 追新的行业热点：比如，短视频火了，要第一时间支持短视频广告；移动直播热了，又要第一时间联合平台进行广告服务。
- 追新的平台热点：在新兴平台着力拓展主营分发业务的时候，是不太顾得上广告中介业务的，这就给了中间商窗口发展红利期。中间商可以复用其他平台的运营模式，快速绑定平台上的新作者，探索平台的变现可能。

对自媒体来说，我个人并不建议其将与广告中介的合作仅当作变现手段。和领平台的流量分成不同，这种内容型广告是用自己的品牌背书、靠粉丝变现的，与品牌定位不符的广告对自媒体是有损的。

自媒体在承接时，一方面应该筛选出与自身品牌定位相符的广告或分销商品；另一方面可以将之视作一次变现价值的实验，关注商业转化率。如果商业转化率到达某个临界点的话，或许意味着应该加快商业化变现的脚步了。

自营电商：隔行如隔山

不甘于被媒体属性束缚的自媒体有更大的愿景，它们开始进一步深挖自己的流量价值和粉丝价值、提供附加服务，从而引出了流量变现的第三个阶段：自营电商。对一些中小规模自媒体来说，可以通过如淘宝客等方式来进行商品导购转化率的实验。

简单地说，“平台分成→自营广告→自营电商”的进程可以理解为：

- 委托平台卖广告
- 自己来卖广告
- 给自己卖广告

广告主向自媒体投放广告的目的除了精准人群的流量曝光之外，也希望能够选择与自己气质相似的自媒体共享品牌形象，并从最终的核销层面获得超过广告投入的收益。当自媒体有能力给自己打广告时，需要的是完成自己体系内的流量变现闭环，即获得比单纯自营广告更大的收益。之所以仍然将其归类到流量变现，是因为内容只是这类自媒体获得流量的方式，最后实现变现的还是实物或虚拟产品。

2016年，内容电商的概念很热，很多头部自媒体都涌了上去。

援引2016年9月新榜发布的《内容电商研究报告》：在近三个月新榜指数800分以上的5269个公众号中，有718个在做电商，即差不多每7个大号中就有一个在试水内容电商。

自媒体电商问卷

想趟自营电商这摊浑水，各自媒体首先要回答以下几个问题：

1. 内容与商品变现紧密度是怎样的？

客观上，不同的内容品类与商品变现的紧密程度是不一样的，有的品类离钱更近，有的离钱更远。你的内容品类，适合卖货吗？如果适合的话，适合卖哪些货？

从新榜和有赞共同提供的销售数据来看，文化类与育儿类公众号的内容转化效率最高。在有赞平台开店的头部大号中，文化类与育儿类的店铺数量仅占总店铺数量的24%，却斩获了71%的销售额，相当于1/4的店铺卖出了3/4的销量。而店铺数量较多的民生类与时事类大号，销售额却只占总销售额的1%。

内容与商品的关联性、商品本身的消费特点共同影响用户的消费决策，不恰当的选品往往会影响销售额。比如，各种自媒体都会尝试销售自身的周边产品（鼠标垫、纸袋、水杯之类），这类商品只能算用户的情怀充值，复购率较低。而一些

标品，如数码3C产品 [1] 、婴儿奶粉、纸尿裤等，用户更容易前往已有的电商平台消费，转化率较低。

2. 短期收益预估如何？

自媒体销售商品，多用自己的内容进行商品内容的推送和介绍。因此，对打算尝试商品销售的自媒体来说，需要意识到这个推送位置本身是有价值的，应当比较商品内容与承接广告二者间收益的不同。

- 商品内容利润＝销量 ×利润 - 成本＝（阅读量×购买转化率）×（客单价×毛利率）-（内容运营成本＋电商运营成本）
- 广告内容利润＝ 收入 - 成本 = 广告收入 - 内容运营成本

假如你的公众号已经能够承接单条6万元的广告了，如果销售客单价为100元、毛利率为30%的商品，就需要销售2000件才能在收入层持平。值得注意的是，在成本层，自媒体团队在实操过程中往往会忽视对电商运营成本的核算。比如售前咨询、售中资询、售后服务，这几项耗费人力的事情基本与销售量保持线性增长关系。

此外，上述公式只比较了单次销售的情况，自媒体如果缺

乏持续稳定的SKU（库存量单位）建设，那么商品销售就会成“三天打鱼，两天晒网”，无法形成规模化效果。

通常，我们建议自媒体以类似淘宝客、京东客的方式切入，自己筛选、分发与淘宝、京东对应的商品，以降低电商部分的运营成本，先进行选品准确率、商品销售转化率的打磨与评估。经过多次销售行为，确认商品收入能与广告收入持平后，再去触探电商运营、销售链建设等复杂度更高的环节，试图在利润环节与广告持平。

绝大多数自媒体涉水电商短期内获得的收益无法与广告内容收益相比。想要触达可能性更大的电商属性，这份愿景或许是支持自媒体不断探索的动力。

3. 电商团队的预估效果如何？

在验证了自身的商品变现能力后，自媒体团队就可以正式下场做电商了，从电商选品到电商内容编写再到售后服务，组建自己的电商团队。

首先要从选品环节切入，符合自身的调性是第一标准。自媒体需要从自身的粉丝构成和账号定位出发来进行商品的筛选，不能仅追求高大上，更要追求符合调性的购买转化。以科技自媒体“差评”为例，在其电商运营早期，选择了很多高大上的黑科技产品，不仅购买转化率不高，用户的反馈也逐渐变得

冷淡。直到一款单价为40元的创意商品Sticker销售2000多件，才促使差评重新审视自己的选品策略。结合粉丝“18～25岁、70%为男性、集中在学生群体、多为互联网从业者和对科技感兴趣的人、消费力不强、相对理性”的特点，差评将自己的选品策略调整为新奇但客单价不高的“小黑科技”产品。赚钱当然也是另一大目标，但不刻意追求大品牌，以自己的内容品牌给产品价值背书，“小众、非标、贴牌”成了大量自媒体的选择。

选对了商品，还需要借助包装对内容进行推荐。同内容要追求点击率、分享指标一样，商品推介内容同样要求制作精良，考核阅读指标和购买转化指标。一条CEO徐沪生就非常推崇宜家团队以精美的装修设计向用户传达感性认知的工作方式。一条的工作方式亦然，有专人负责商品推介视频的策划和拍摄，甚至会介入商品的选品环节。如果某个商品不符合一条的素材要求，就会被放弃。

商品选对了，内容包装好了，销售额上去了，自媒体就又迎来了甜蜜的新烦恼：客服电话响个不停，售前的商品咨询，售后的问题解答，物流配送的及时性等各种细枝末节的事情。哪个环节掉链子都会影响消费者的体验，造成掉粉。

4. 供应链稳定度的预估效果如何？

对电商运营团队来说，进一步的压力来自供应链的稳定程

度：生产、仓储、物流，哪个环节跟不上都够自媒体头疼的。如果说物流问题影响用户体验，仓储问题加剧成本负担，那么生产质量问题就扎扎实实地影响了品牌的公信力。

在过去的一两年里，海淘的热潮逐步兴起。由于海淘商品对国内市场来说多为新兴品牌，比价性较弱，故有很多自媒体倾向于选择同跨境供应商合作销售海淘产品。随之而来的是对供应商审查不足导致的假货问题，母婴、美妆类产品都是重灾区。

鉴于供应链的复杂度，有些自媒体选择“做轻”的模式，即只做选品，通过与供应商合作，交由供应商仓储发货。徐沪生认为，一碰库存就是巨大的成本。尽管一条2017年的销售额预估为10亿元，但目前公司仍然选择无库存的电商平台模式，库存、发货均由供应商负责。

典型案例：年糕妈妈

出于对母婴教育市场的兴趣，我很早就开始关注亲子类自媒体，下面也会以“年糕妈妈”为例进行阐述。

2014年，喜获宝宝的李丹阳开始了“年糕妈妈”微信公众号的创作。这时的她，更多是由兴趣驱动，做着基于自身知识、育儿经验的分享。

2015年，年糕妈妈开始出现超过10万阅读量的内容，粉

丝数超过10万，算是实现了第一个突破。进而，年糕妈妈组建了自己的编辑团队，开始将一个兴趣爱好正式变成项目来运营。

深耕后的项目开始有针对性地覆盖教育、辅食、陪玩、孕产等不同垂直领域，内容更有体系化和专业度，面向特定受众以内容矩阵的方式来抢占市场。在这一阶段，年糕妈妈已然是一个职业运作的亲子自媒体。

2016年，“糕妈优选”电商平台上线，进入母婴电商领域，首次团购销售额超过百万元。在这一年里，年糕妈妈把内容和电商更紧密地结合，融资以扩充规模，用内容影响消费决策，依托独代非标商品获得更高毛利。

如果你已为人父母，就会知道，纸尿裤、奶粉、童书、玩具等都算得上刚需。各大母婴类电商都在把此类刚需型产品作为引流点赔本赚吆喝。对自媒体来说，在标品市场与大平台竞争无疑于以卵击石，只有经营非标品才能够避免用户的比价冲动，更容易促成交易、收获利益。

但非标品也存在一个用户认知的问题：用户为什么要花更多的钱购买一个对主流消费市场而言并不熟悉的品牌呢？自媒体的内容教育，在这个过程中扮演了关键角色。

以防晒产品的销售为例，年糕妈妈通过内容向用户介绍美

国儿科学会建议的儿童防晒的注意事项、儿童防晒霜的成分。在知识介绍后再开始商品的宣导，并对跨境主流产品进行测评，最后落点是自家销售的产品。

为什么只测评跨境产品？其根源就在于对跨境商品来说，保真是第一，敏感度、价格已经不那么重要了。除了做海淘非标品之外，700万的粉丝量也让年糕妈妈有底气做独代的生意，进一步扩大毛利。

成长之路难免坎坷，年糕妈妈也遭遇过“假货风波”。2016年，用户反馈购买的商品有假货嫌疑，与客服沟通过程中遭遇了种种不顺，最终演变成微博上一个专门的话题“年糕妈妈卖假货”，成了年糕妈妈成立以来的第一次公关风波。站在从业者的角度，我个人更倾向于相信这是由于对供应链把控不足以及客服体系跟不上导致的。在商言商，供应链和客服体系是自媒体人进化到内容电商需要补习的一课。

2017年，年糕妈妈的商业化之路步伐更快，除了不断完善实物类商品销售外，还上线了“育儿课堂”这样的知识付费产品。打造专家化团队、树立权威IP形象，年糕妈妈开始讲一个更大的故事。

很荣幸，2016年年底，我参加了有赞4周年的年会，会上年糕妈妈CEO林威的发言让我得以一窥自媒体自营电商的门道。彼时，年糕妈妈的月销售额就已经稳定在5000万元左右

了。援引年糕妈妈的自我剖析：

- 坚持良好的内容，其前50万粉丝是零推广费用，靠内容传播获取来的。
- 稳定的供应链，包括：自建仓库，全国有三个货仓，双十一支持发货60万件；建立跨境货仓，覆盖海淘业务。
- 1/3的独家代理产品，保证利润率。
- 自己造货，加速品牌变现过程，保证利润率。

会上，林威分享了这样一种观点：对于“内容电商”这一概念，电商才是主体，其商业逻辑的链条是“内容→粉丝→复购”的传导过程。在整个过程中，内容主要起到推广的作用，即内容好坏影响的只是电商的获客成本。但对电商来说，最看中的并非获客成本而是复购率。在复购环节，影响用户决策的就不仅仅是内容的阅读体验了，电商本身的下单、物流、客服体系同样重要。自媒体做电商最大的阻碍，不是缺懂内容、懂粉丝的文化人，而是缺懂货品、懂供应链的生意人。

---

[1] “3C产品”是计算机类（Computer）、通信类（Communication）和消费类（Consumer）电子产品的合称，也被称为“信息家电”。——编者注

本书由“ePUBw.COM”整理，ePUBw.COM 提供最新最全的优质电子书下载！！！

内容付费

在最近几次同自媒体的沟通中，“内容付费”这个关键词出现的频率越来越高。当越来越多的自媒体高喊让内容不再只是流量，让内容回归其价值本质的时候，内容付费是否迎来了自己的春天，能够真正成为一门好生意呢？

不能免俗地回顾一下自2016年开始，在内容行业影响比较大的几个内容付费玩家的玩法。

罗辑思维&得到

尽管罗振宇早在2013年就因为会员付费成为自媒体行业中的潮流引导者，但真正以“内容付费”为核心引发讨论的还是其旗下“得到”App于2016年6月5日推出的年费为199元的李翔商业内参。由马云、柳传志等企业家加持的李翔商业内参一经面市，就销售万份，营收破千万元。

在李翔商业内参之外，得到坚定不移地践行着“知识付费”的理念，上线了多位头部大咖（如王煜全、吴军等）的包年制专栏产品。其中，李笑来专栏成了销售冠军，共计收到12万份付费订阅、创造了2500万元营收。除了专栏产品之外，得到还提供了内容主题更为分散的付费音频和电子书内容。

2017年2月21日，脱不花在一次直播中披露了得到的数据：目前，得到App的总用户有529万，其中每天的活跃用户数约为42万，总订阅量为130万，总订阅人数近80万。据此估计，得到的营收规模在1.5亿元左右。

得到主做头部的模式，被大家比作知识付费里的奢侈品店。以罗振宇主IP为基础的巨大流量，与平台上邀约的各个头部大咖的自有粉丝流量叠加，共同加热了“内容付费”这一概念，并树立了得到内容付费第一品牌的用户认知。

喜马拉雅

尽管喜马拉雅在国内音频市场是当之无愧的领头羊，但不赚钱始终是它最大的痛点。2016年6月6日，喜马拉雅携手马东上线了知识型音频节目“好好说话”。该节目凭借着米未传媒“奇葩说”节目打下的良好粉丝基础，销售逾10万份，营收破千万元。

“好好说话”的商业化成功，让喜马拉雅在音频付费的道路上一路狂飙，快速上架了多款音频付费节目。在音频课程建设上，喜马拉雅采取的方式是引入许多垂直行业的头部IP，与之合作建设课程，并辅以课程培训群等方式来优化用户体验。由此，我们更多地将其看作天猫模式。

2016年12月3日，喜马拉雅举办了首届“123知识狂欢

节”，最后实现交易额5088万元。我们参考电商网站年糕妈妈单日营收额和全年营收额的比例，喜马拉雅在内容付费上的营收额预估在5亿元以上。内容付费和车载场景或许是喜马拉雅未来的两个营收点。

知乎

2016年的知乎在数字出版之外，对内容变现做了更多的探索，典型的两款产品是值乎和知乎Live（直播）。前者被视作狙击分答的产品，后者则是类似微信群分享课程的“一对多”产品。援引官方披露的数据，截至2017年年初，知乎Live已经举行了超过2900场直播，讲者的平均时薪超过了11000元，有超过269万人次参与过直播，用户复购率达到了43%。

得益于本身的内容创作生态，知乎沉淀了一批优质的中长尾创作者，知乎live的内容切入点也或多或少地倾向于自家的创作者。与得到和喜马拉雅相比，知乎Live的产品，似乎更像是淘宝模式：人人都可以是生产者和销售者。

2017年5月17日，在知乎Live上线一周年之际，知乎App上线“知识市场”，汇集了知乎Live、知乎书店、付费咨询三种主要的付费服务形态。这或许意味着，跻身独角兽公司 [1] 的知乎除了广告变现外，要把内容付费当作一门主要的生意来做了。

得到（奢侈品店模式）、喜马拉雅（天猫模式）与知乎（淘宝模式），共同构成了内容付费平台级典型案例。它们搭建了内容市场，提供了内容商品的上架、销售与售后全流程支持，并或多或少地介入了内容的采购与制作环节。

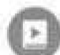
●●●●● 中国移动  上午9:51 53%

发现

关于衬衫大多数人
不知道的15条小知识
戳此查看




电子书


每天听本书


商城


系列


全部

每天听本书 查看全部 >



《三体Ⅲ：死神永生（下）》| 怀… 今日新上

怀沙 讲述 时长：29分4秒

“中国当代科幻第一人”刘慈欣代表作，一切都是未知的宇宙里，人类向死而生。

¥4.99 VIP价 ¥ 0 购买

免费专区 ▷ 连续播放

 李翔知识内参 查看全部 >

颠覆性变革的三个基础条件

柳传志的做事方法：布局和余量

团队管理：慢下来才有互动效果
02:04 

 发现  今日学习  随时听  已购  我的

图13–1 得到App手机端界面图

中国移动 上午9:51 53%

搜索知乎内容

知乎 Live

主讲人：Sunlau

租来的房子
也能装出「家」的味道

Live
互动分享会

书店
精品电子书

付费咨询
线上语音解答

每日新知

健康：如何通过嚼口香糖保护牙齿 02:25

心理：恋人间的吸引有几重境界？ 02:28

博闻：盲肠会随进化消失吗？ 02:30

▶ 免费播放全部

查看更多 ›

今日优惠

精选内容 • 限时特价

距结束 14 : 08 : 47

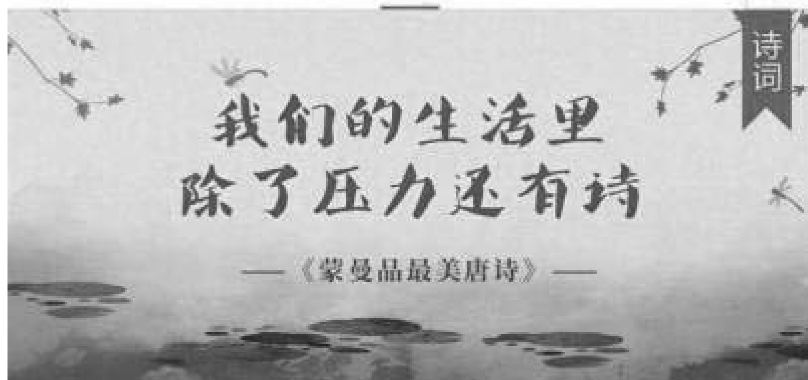


 首页
 想法
 市场
 消息
 更多

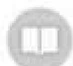
图13–2 知乎手机端界面图





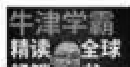














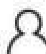


图13–3 喜马拉雅手机端界面图

得到、知乎等公司的“快”，与传统的内容分发平台微博、微信的“慢”形成了鲜明的对比。

微博的付费订阅、付费阅读已经不是什么新鲜事儿了。在2016年年末，微博才终于推出了自己的付费问答产品。而对于这款产品的定位，微博CEO王高飞给出的答案是：微博问答对于微博，不是子产品，也不会有独立入口，而是自媒体赋能工具之一，跟打赏和付费订阅一样。

微信的步调更是一如既往的审慎。2017年2月15日，马化腾在Keso（洪波，著名独立评论人）的朋友圈下留言，阐明微信公众号的付费阅读机制即将上线。但直到写下本文时，该功能仍未上线。

为什么慢？因为心中有底气：平台方自身的盈利状况良好，不需要通过内容变现；由于拥有足够的粉丝流量和成熟的变现方式，核心创作群体们的黏性和留存很好。

对于微博，摘录“财经天下”对新浪微博的采访：

> 先在微博上变成网红、继而向淘宝店铺导流的电商模式，是大号、中号们最赚钱的手段。这种自媒体可能跟大家熟知的媒体或微信公号不一样，是商业类的自媒体。这些网红不但拥有强大的粉丝，还会自主地生产内容。而他们的粉丝也更愿意为他们买单，甚至还自豪地称之为“爱的

供养”。

对于微信，通过微信群承载粉丝社群互动，通过“阅读原文”外链引流电商的行为屡见不鲜。援引新榜《中国微信500强月报（2017.01）》中对于阅读原文的使用情况：有10%的链接导向了电商，2%导向了下载。

因此，内容付费对微博、微信来说，只是其成熟的粉丝经济中的一个应用而已。很多头部自媒体已经开始通过各种方式进行内容变现了，比如吴晓波频道、年糕妈妈等。

以吴晓波频道为例，其构建了三级会员体系：频道新锐会员（每年180元，195000人开通）、思想食堂会员（每年4800元，1588人开通）、企投会会员（每年69800元，399人开通），对应的权限分别为线上付费内容、线下讲座以及类MBA（工商管理硕士）式的课程。会员体系整体开通人数突破20万。



图13–4 吴晓波频道截图

与此同时，为服务于这些头部自媒体，也衍生出了一些相应的产品，以下是一些参考案例：

- 小鹅通：最初由头头是道基金（吴晓波与曹国熊为发起人）投资，为微信公众号提供内容付费、用户管理、社群运营等SaaS（软件即服务）服务。已入驻吴晓波频道、张德芬空间、十点读书、局座召忠等数百位自媒体头部账号。近日，小鹅通完成了3000万元的A轮融资，投资方为喜马拉雅。

- 知识星球：相当于一个付费版的读者社区，用户购买的是可以参与更多探讨的会员资格，能够在共同关注的自媒体或用户间进行更多的交流，有caoz（曹政）、小道消息等入驻。

在基于粉丝的内容付费生态相对繁荣的情况下，是否还需要平台方，已经显得不那么重要了。

可以预见的是：作为自媒体变现的一种方式，内容付费会逐步出现在越来越多的内容平台上，成为内容平台的标配。尽管各家平台对其内容付费产品的支持力度，会因自身内容来源分布、营收方式不同而有强有弱，但内容付费这种形态，预计会在整体市场的共同投入下，快速繁荣。

内容付费之火，或将燎原。

缘何付费，规模几何

下面，我们试图探讨内容付费为什么是可以成立的，用户是怎样做出付费决策的。

消费背景

内容付费并不是什么新鲜事儿。只是在最近一两年里，付

费的渠道和内容的载体开始往线上迁移。我们通过盘点现有的内容付费形态就会发现，它们往往能够对应到已有的线下模式。比如：

得到、喜马拉雅和36氪的预付费专栏创作模式，可以类比期刊的线上版。如果非得纠结专栏是个人创作内容的话，那或许可以膜拜一下郑渊洁先生独自创作运营数十年的《童话大王》。

分答、值乎的模式是将线下的付费咨询搬到了线上。一些医疗领域的线上问诊，更具有付费典型性。

知乎Live、群分享课程可以类比线下付费培训、讲座的线上版本。

那么，为什么搬到线上后，内容付费可以有更大的想象空间？

第一，当然是服务体验的提升。

对图书杂志有购买意愿，对线下的讲座培训有购买意愿，对电影、话剧等娱乐形式有消费意愿，这充分说明用户对内容是有消费需求的。当内容从实体型或线下型转变为线上型的时候，其使用体验是有延展和增益的。

以知乎Live作为样板分析，每次通常为1小时左右的线上

直播课程，分为主讲人播讲的内容主题、在线用户互动提问两部分。在直播结束之后，感兴趣的用户还可以通过付费咨询的方式与主讲人进一步沟通。

这种形态的创作和传播成本更低，具有更好的时效性：新的案例、新的观点可以通过直播的形式更快地传递出来。对比书籍、杂志的形态，当书籍、杂志付诸印刷的时候，谈论的可能已经是两个月前的旧事了。

同时，由于引入了线上互动和线下咨询的环节，使得用户的标准化信息获取需求在通过直播主体达成之余，非标准化的个体咨询需求也可以通过点对点咨询完成。

线上传播不受时空限制的特点，除了对内容消费者是利好外，也可以帮助小众领域的创作者聚集更多的潜在消费者，从而促进其生产，最终带来内容消费品的品类和数量繁荣。

第二，小额付费、版权意识增强客观上完成了对用户群体的教育。

基于微信、支付宝的在线支付方式的便捷性是远胜往昔的，用户越来越习惯使用手机进行小额的、及时性的消费。随着版权市场的日趋规范，音乐会员、视频会员的付费规模日趋增长，对用户也是有正面教育意义的。

消费需求、消费途径、内容产品，三者已经构成了一条完

整的通路，剩下的问题只是不同类别的内容消费能够做多大了。

消费意图

用户对内容的消费意图，分为功利性诉求和非功利性诉求。功利性诉求在于解决自己的问题，最突出的是各类型应试需求的课程（如考研培训、公务员考试培训等）；非功利性诉求是通过内容消费来满足自己的好奇心或达到对某个领域的初步了解甚至深度学习的目的。

如果要描述功利性和非功利性诉求区别的话，我想，可能是前者需要的是通过考试和学知识，后者需要的是长见识。

通过考试，是刚需且反人性的，大部分人的应试准备过程都需要被督促。四六级培训、考研培训、公务员考试培训、专业考试培训等是功利性消费的典型，有着明确的消费预期和结果量化方式。

学知识，是刚需但需要自驱。比如，新手父母学习非暴力沟通、职场人士补习英语等，都是为了解决真实存在的问题。但是否能够学完，学完之后是解决了问题还是仅仅解决了焦虑感，就因人而异了。比如，我曾经付费英语流利说的半年课程，在对应的百人学员群里，能够完成整个学期课程的人，只有百分之三四的样子。

长见识，是非刚需，甚至可以归类为泛娱乐化消费的范畴。比如，听郎朗讲如何弹钢琴、听从业者讲OFO（线上线下融合）市场运营，更多是为开眼界；消费李翔商业周刊，更多是获取谈资。在这种场景下，用户不会有明确的结果预期（即需要收获什么），只要“大开眼界＋不明觉厉”就已经很好了。

知识与见识对比起来，前者需要沉淀，是相对结构化和稳定的；后者个体主观性更强，迭代更为快速。

如果抱着学知识的目的去长见识，结果往往会令人大失所望、心存落差。我们分析知乎Live的课程评价，常常见到两极分化的情况，无论怎样的课程都会有人抱怨太浅、没有干货。图13–5是郎朗的Live差评截图，作为一场面向12000人的直播必然是普适性的，以干货来要求它或许是预期错位了。









图13–5 郎朗直播差评截图

消费决策

进一步地，用户是如何做出消费决策的呢？

核心还是一个字：值，即用户对内容的价值预期与内容价格相匹配。

价值预期的构成可以有很多种，一种常见的观点是价值稀缺论。这种观点称：当信息爆炸、免费内容泛滥的时候，人们愿意通过付费的方式让自己更值钱的精力消费在更优质的内容上。

然而，就我个人而言，并不太同意这种说法。以李翔商业内参的热销为例，一个朋友的评论很精辟："你都不愿意为李翔及其一众编辑精心编写的杂志付费，怎么就愿意为他一个人产出的每天豆腐块点评付费呢？"这或许说明了至少在这件事情上，用户的价值预期判断不仅仅是基于内容本身的价值的。

价值预期的构成可以有很多种，比如内容本身的价值、主讲人的品牌价值、平台的品牌背书等。

内容本身的价值，最直接的体现是用户感知到内容消费是对自己有帮助的，可以迅速转化为产出。参考企鹅智库的调查数据：能提高工作效率或收入的知识和经验最被用户认可，有63.3%的人有意愿付费；有关职业与学业的发展建议，也达到将近四成的付费意愿。因而，内容产品"功利性"的包装成了不少知识出品方的选择：

- "好好说话"的定位是："从沟通、说服、辩论、演说到谈判，教给用户一整套应付生活场景需求的话术。"
- 李笑来的"通往财富自由之路"，其定位更是直接体现

在了专栏标题上。其主讲的主题为“一小时建立终生受用的阅读操作系统”的知乎Live，更是吸引了超过12万人的参与，刷新了知乎Live单场最多参与人数的纪录。

“功利化”的内容包装更容易促成付费，我们参看知乎上参与人数较多的直播，大多主打实用性的特点。付费内容能否解决具体问题依赖讲者和听者的共同作用，但一定能够被解决的是听者的自我焦虑感。

主讲人的品牌价值，市面上销售量较高的内容大多都在此列。罗永浩的“罗永浩的创业课”、李开复的“十堂人工智能课”、高晓松的“矮大紧指北”都可以说是主讲人自带流量光环的典范。我和身边的朋友，都购买了罗永浩的付费专栏，权当为信仰充值了。在这种情况下，用户消费的其实不是内容，而是人。与其说是内容型消费，不如说是粉丝型消费。

平台的品牌背书，如李翔、王煜全的专栏销售，尽管他们已经是各自行业中的翘楚，但是在大众内容消费市场的知名度并不高。马云、柳传志等大咖对李翔的背书，罗振宇对王煜全的力荐，以及“得到”早期对罗辑思维粉丝流量的依赖，都构成了对这些内容价值预期的背书。于是，内容产品的介绍信息变得越来越像书籍的腰封，人们通过精美的装帧、销售排行榜和封面上一个又一个名头来做出是否消费的判断。

规模几何

对于内容付费市场的规模，计算方式有多种。

第一种，与图书市场类似。图书消费同样有解决问题型和消遣型的诉求，与内容消费类似。此外，国人对于书籍的消费通常是购买而不是借阅，因而更有借鉴意义。2016年，我国图书市场规模为690亿元，扣除教辅（占比17%）和少儿读物（占比20%），剩余部分的市场规模约434亿元。

第二种，与视频网站付费会员类似。视频网站的会员付费行为大多是被独有或头部内容所驱动，从这个角度看与内容付费有一定的相似性。援引相关统计数据，2016年全网视频付费用户数约为7500万，规模达到96亿元；预计2017年全年付费用户将超过1亿，在ARPU（每用户平均收入）值保持不变的情况下，规模将达到128亿元。目前，大家普遍对视频付费会员市场保持乐观态度，预计未来还会有进一步的增长。

无论是哪种计算方式，内容付费都应该是百亿元规模，并会保持一段时间的高速增长。

目前，得到、喜马拉雅、知乎三家的年营收额加起来乐观预估在30亿元以内；独立的头部IP，如樊登读书会、吴晓波频道等营收规模都刚刚过亿元；其他的内容IP营收多在千万元、百万元量级。

内容付费市场，或大有可为！

内容付费平台展望

我更愿意将内容付费市场与图书市场相比，平台方就像出版社，头部自媒体像畅销书作家，垂直领域头部自媒体像专业领域作家。

对平台而言，内容付费可以做得很轻，也可以做得很重。

做得轻，是只做产品功能上的对标。常见的功能无非是付费问答、付费阅读、付费课程、会员制等，将各种已有内容进行包装，并加以付费门槛和权限控制。基于粉丝关系分发的平台，都可以将内容付费作为赋能创作者的一种手段。而那些不基于粉丝关系分发的平台，则可以将内容付费作为自身流量变现的一种途径，通过分销热门课程来赚取分佣。

做得重，在当下的时间点上，就不可避免地需要介入上下游产业。

上游是解决内容的问题。内容背后站着人，想要解决内容的问题，就需要明确将人摆在什么位置。

一种方式，是做标准化的课程体系，类似英语流利说、VIPKID（一家在线少儿英语教育公司）。在这种场景下，人的

品牌是弱化的，平台方更追求内容的实用价值和服务配套体系。比如，产出一系列关于新生儿护理的内容，在领域知识已经非常标准化的情况下，由谁来播讲已经变得不重要了。我们也许可以将其类比为印刷出版工具书（字典、地图、教材）。

另一种方式，是做非标准化的分享课程，比如吴晓波频道、keso专栏等。在这种场景下，人的品牌无疑是占主导地位的，无论是内容的选题、创作、更新频率等，均收归创作者决定。毕竟，对头部自媒体来说，单单靠广告收入就已经赚得盆满钵满了。付费内容既可以视作一条创收道路，又可以视作粉丝贡献度筛选途径。

援引《第一财经周刊》的报道：在keso与36氪的合作中，“平台方要求专栏是独家的，keso则坚持个人微信公众号必须发；分成比例上平台也模模糊糊地说，按扣除营销等费用后的收益分成，keso则坚持按总收入谈一个具体数字”。由此，不难看出平台与强IP的合作中，谁更具有话语权。

由于缺乏足够的议价权，平台更多地变成了大IP的流量分发渠道。以李开复的“十堂人工智能课”为例，就同时分发在了蜻蜓、喜马拉雅等多个平台上。李笑来一边在各处开专栏，一边又搭建起了自己的音频付费平台。我们也许可以将这种情况看作某个出版社签下了畅销书作家的大作，尽管在合作中，出版社也未见得能够分得多少利益，但是本着赚吆喝的心态还是

会力争此类内容的合作。

而上述两种方式的中和，走平台深度定制、指导生产的方式，或许是未来内容繁荣的切入点。通过对比图书市场，我们不难发现，现在的内容付费市场是不完备的，只有畅销书高声叫卖，缺乏各个垂直领域书籍的覆盖。如果我们预期未来市场消费繁荣的话，必将倒推到内容产品的数量繁荣上。

从已有的案例来看，垂直领域的头部作者经由平台背书和挖掘后，具有变现价值进一步放大的潜力。比如曾登上百家讲坛的中央民族大学历史系副教授蒙曼在喜马拉雅开设的“蒙曼品最美唐诗”，预估就有百万元销售额。各行各业的腰部IP，具有更高的自我传播意愿和平台合作度，结合平台的数据分析，有望生产出市场与口碑双赢的付费内容产品。

对腰部的自媒体而言，已经有一些在自己的社群内进行付费尝试了。只是，十万规模的粉丝、百分之几的付费率、百元的客单价，这笔账怎么算也只有几十万元的营收规模。在内容付费的红利点上，积极寻求与平台的合作、借助平台的力量同时扩充营收规模与粉丝规模，或许是一种解决方案。

这很类似于各个出版社在构建自己的出版物体系时，持续与新的创作者合作以覆盖各种垂直领域和人群，并试图从中挖掘出新的畅销书作家。作家与出版社在合作的过程中逐步实现互相选择与互惠共赢。

下游旨在解决用户端的消费体验问题。

付费内容大多以音频课程的形式开展，交互形态还比较初级、收听体验差、内容同质化、水化、缺乏服务体系和有效的效果评估是目前普遍存在的问题。如何规范服务，让非标品也能够拥有标准化的售前、售中、售后环节，或许是内容付费平台建立行业规范的一个机会。

其中，售前环节注重信息匹配的效率和准确度。目前的流量分发方式大多集中在人工运营，以各种横幅广告的方式来推介头部内容。伴随着中长尾内容的数量繁荣，势必需要引入更高效的分发方式，以降低整体分发系统的基尼系数。以得到的方式来分发知乎Live的内容，显然是行不通的。

同时，考虑到付费内容的消费体验与用户接受程度息息相关，如何更好地展示售前信息，把对的内容推荐给对的用户也给售前分发环节提出了更高要求。比如，我的知乎Live推荐中就不时出现“大学生实现经济独立的7大策略”“考研英语过关攻略”之类匹配度极差的产品，受众错误自然毫无点击欲望。对付费产品的个性化分发不应仅以购买转化率作为唯一指标，同时需要考核的还有产品的退款率、用户的评分预估等。

售中环节注重消费体验的优化。以“音频＋图文”为主体，

参与者互动提问已经是市面上的标配了。围绕已有环节，如何保证音频的收听音质、视频课程的画质、控制课程长度、合理安排章节分布等都是可以深挖的产品功能点。

在已有环节之上，如何延展消费体验，比如建立用户间的联动、从单次内容付费开始建立中长线的用户到粉丝的转化、帮助创作者构建影响力和粉丝体系，或许都有可尝试的空间。以图13–6为例，这是喜马拉雅上的一节付费课程，付费用户可以进群后同主播进行进一步的沟通。



图13–6 喜马拉雅付费课程界面截图

在售后环节，需要以标准化的评估方式约束非标准化的内容创作，争取做到既让用户端满意，也让作者端有充分反馈。比如，知乎对直播市场做出的一系列规范措施：首先，面向用户端推出7天无理由退款功能，即用户在知乎上购买直播以及

直播结束后的7天内，如收听语音未超过15条，可以无理由退款；其次，面向作者端升级评价体系和版权服务。通过不断规范平台交易流程，一套更高效的知识市场体系或许值得期待。

---

[1] 独角兽公司：投资界对估值在10亿美元以上，并且创办时间相对较短的公司的称谓。——编者注

本书由“ePUBw.COM”整理，ePUBw.COM 提供最新最全的优质电子书下载！！！

图书在版编目（CIP）数据

内容算法：把内容变成价值的效率系统 /闫泽华著. --北京：中信出版社，2018.4

ISBN 978-7-5086-8527-4

I. ①内… II. ①闫… III. ①网络信息资源 - 资源利用 IV. ①G250.73

中国版本图书馆CIP数据核字（2018）第007771号

内容算法——把内容变成价值的效率系统

著者：闫泽华

出版发行：中信出版集团股份有限公司

（北京市朝阳区惠新东街甲4号富盛大厦2座 邮编100029）

电子书排版：萌芽图文

中信出版社官网：http://www.citicpub.com/

官方微博：http://weibo.com/citicpub

更多好书，尽在中信书院

中信书院：App下载地址https://book.yunpub.cn/ （中信官方数字阅读平台）

微信号：中信书院

本书由“ePUBw.COM”整理，ePUBw.COM 提供最新最全的优质电子书下载！！！